Theoretical Investigation of Cisplatin-Deoxyribonucleic Acid Crosslink Products Using hybrid Molecular Dynamics + Quantum Mechanics Method

YAN, Changqing

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Philosophy

in

Chemistry

The Chinese University of Hong Kong

September 2009

Thesis/Assessment Committee

Professor Jimmy Chai Mei, YU (Chair)

Professor Steve Chik Fun, AU-YEUNG (Thesis Supervisor)

Professor Sherlock Sik Lok, LAM (Committee Member)

MASTER OF PHILOSOPHY (2009) The Chinese University of Hong Kong
(Chemistry)

TITLE: Theoretical Investigation of Cisplatin-Deoxyribonucleic Acid Crosslink Products using Hybrid Molecular Dynamics + Quantum Mechanics Method

AUTHOR: YAN Changqing, B.Sc. (University of Science and Technology of China)

SUPERVISOR: Professor Steve Chik Fun, AU-YEUNG

NUMBER OF PAGES: x, 97

Thesis/Assessment Committee:

Professor Jimmy Chai Mei, YU (Chair)

Professor Steve Chik Fun, AU-YEUNG (Thesis Supervisor)

Professor Sherlock Sik Lok, LAM (Committee Member)

ABSTRACT

The ligating property of DNA has been theoretically investigated from the structural and electronic point of view in this thesis. A series of short segment DNA models in solvent was studied using a novel hybrid Molecular Dynamics + Quantum Mechanics (MD+QM) method. The hybrid method was demonstrated to be highly applicable for the studied systems by comparing the calculated results with the experimental data. Based on the calculated results, a set of empirical rule which locates the two most active sites in DNA sequences leading to postulation of cisplatin-DNA crosslinks was proposed. The rule was extremely successful when applied to the experimental sequences. Then the four experimental cisplatin-DNA adducts were carefully examined structurally and a better understanding of the interaction between cisplatin and DNA was deduced. Some other applications of the calculated results were also briefly discussed in the thesis.

摘要

本论文从结构学和电性质的角度，在理论上研究了脱氧核糖核酸（DNA）的结合性质。一种新型的 MD+QM 杂化理论计算方法被采用，借以研究一系列在溶液中的短序列 DNA 模型。通过计算值和实验值的比较，这种杂化理论计算方法对所研究系统的高度适用性被证实。在计算结果的基础上，提出了一系列经验性的规律来定位 DNA 序列中最高活性的二个位点，借以预测顺二氯二氨基铂与 DNA 的结合产物。这些规律在应用到实验上的 DNA 序列时取得了非常大的成功。本论文随后在结构上仔细检视了四种实验上发现的顺二氯二氨基铂与 DNA 的加成物，取得了对顺二氯二氨基铂与 DNA 的相互作用更深的理解。本论文还简要讨论了计算结果的一些其它应用。

ACKNOWLEDGMENTS

I would like to express my sincere thanks to my supervisor Professor Steve Chik Fun, Au-Yeung for his kind guidance, invaluable advices and most importantly great encouragement and support.

Special thanks are given to Cheung Chung Hong, Yang Lifeng, Chen Xiangfeng and Chen Chenwen for their technical supports in computing. Without them I can't complete the research project in time. Thanks are also given to the labmates in LG52B, Science Center in The Chinese University of Hong Kong (CUHK) for their support and encouragement.

Finally I would thank my friends and family members for their care, patience and unselfish love. They are the greatest power for me to accomplish this project.

LIST OF ABBREVIATIONS

DFT	Density Functional Theory
DNA	Deoxyribonucleic Acid
FMO	Fragment Molecular Orbital
HF	Hartree Fock
HOMO	Highest Occupied Molecular Orbital
HOMO-1	Next Highest Occupied Molecular Orbital
IP	Ionization Potential
LUMO	Lowest Unoccupied Molecular Orbital
MD	Molecular Dynamics
MM	Molecular Mechanics
MO	Molecular Orbital
NBO	Natural Bond Orbital
NDB	Nucleic Acid Database
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
PME	Particle Mesh Ewald
QM	Quantum Mechanics
RHF	Restricted Hatree Fock
RMSD	Root Mean Square Deviation
SCF	Self-Consistent Field
SPE	Single Point Energy

TABLE OF CONTENTS

(cont'd)

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

BACKGROUND INFORMATION

1.1 Introduction

Deoxyribonucleic Acid (abbreviated DNA) is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms and some viruses. It was first isolated by the Swiss physician Friedrich Miescher who, in 1869, discovered a microscopic substance in the pus of discarded surgical bandages. As it resided in the nuclei of cells, he called it "nuclein" (1). In 1919 Phoebus Levene identified the base, sugar and phosphate nucleotide unit and suggested that DNA is consisted of a string of nucleotide units linked together through the phosphate groups (2). However, Levene thought the chain was short and the bases repeated in a fixed order. In 1937 William Astbury produced the first X-ray diffraction pattern which showed that DNA had a regular structure (3). But until 1953, what is now accepted as the first accurate model of DNA structure was not suggested in the journal *Nature* by James D. Watson and Francis Crick based on X-ray diffraction images taken by Rosalind Franklin and the information that the bases were paired (4). Experimental evidence for Watson and Crick's model were published in a series of five articles in the

same issue of *Nature*. Due to the great contribution, Watson, Crick and Wilkins jointly received the Nobel Prize in Physiology or Medicine in 1962.

Nowadays it's known that DNA structures are more complicated and multiform. It bears a variety of secondary and tertiary structures such as hairpins, bulges, branched junctions, two-dimensional lattices, three-dimensional cubes and octahedrons (5, 6) depending upon the conditions. It's also known that many diseases resulted from malfunction, mutation or damage of DNA such as Downs syndrome, phenylketonuria, hemophilia, skin cancer and so on. To get insights into biological functions of nucleic acids and a deeper understanding of those diseases, the structure characteristics and some chemical properties of DNA were investigated in this thesis.

1.2 Deoxyribonucleic Acid

DNA is composed by two long strands of simple units called nucleotides. The two strands run in opposite directions to each other and are therefore anti-parallel (Figure 1.2). A nucleotide consists of a deoxyribose sugar, a heterocycle base and a phosphate group (Figure 1.1). The sugar part and the phosphate group part, connected by phosphodiester bonds, are usually called "Backbone of DNA". The heterocycle base (Figure 1.3) is connected to C1' position of the sugar ring and the complementary bases are held together by hydrogen bonds to form base pairs, whereby adenine (A) pairs with thymine (T) and

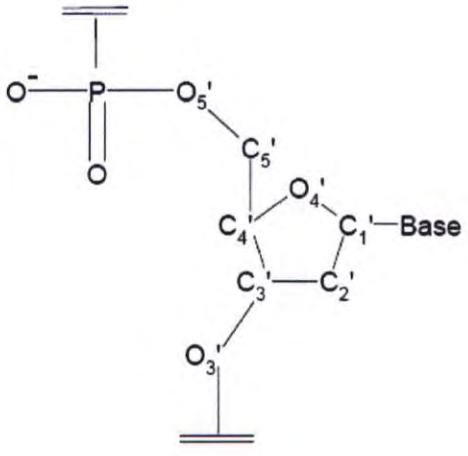


Figure 1.1 Chemical Structure of the Nucleotide

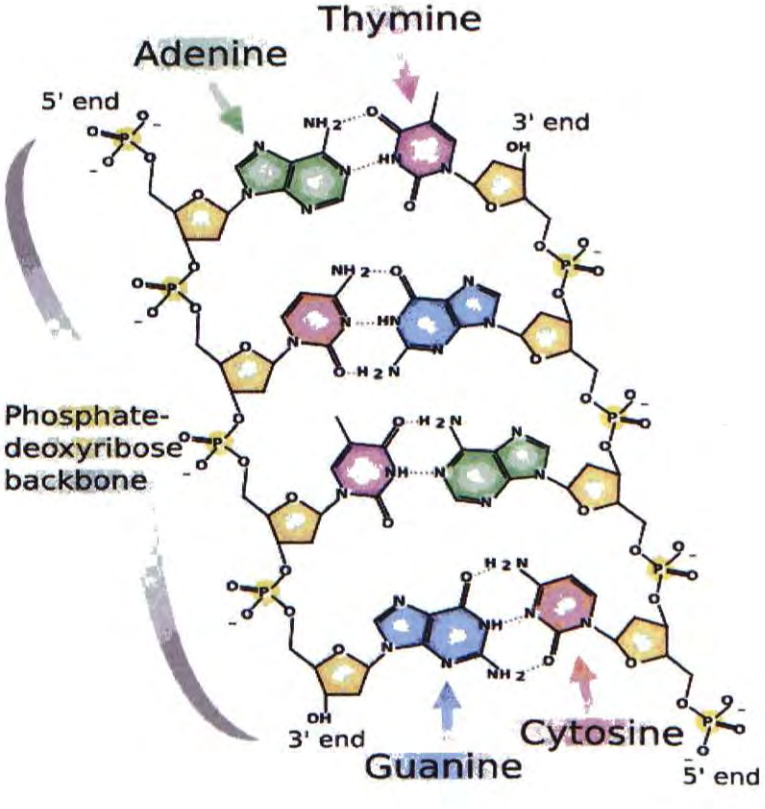


Figure 1.2 Chemical Structure of DNA (Hydrogen Bonds are Shown as Dotted Lines)

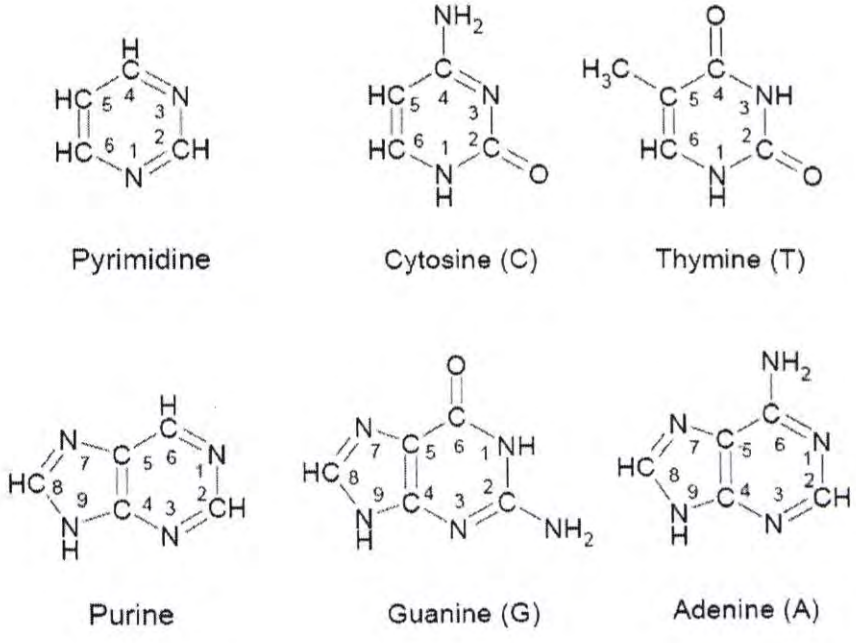


Figure 1.3 Nitrogeneous Bases of DNA

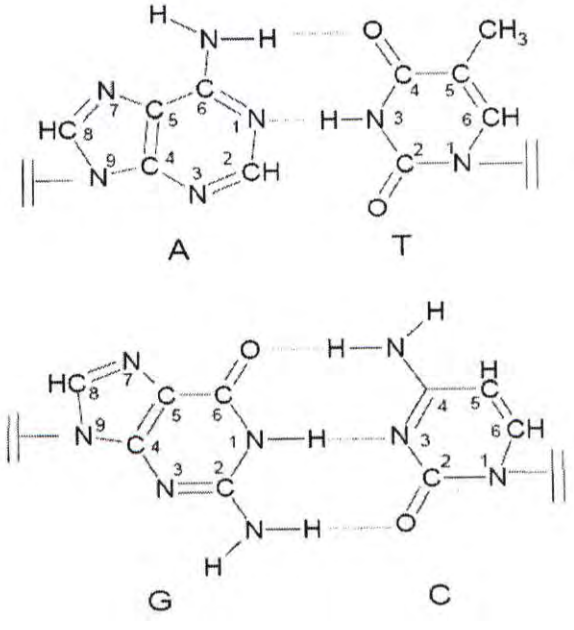


Figure 1.4 Base-Pairs of DNA (Hydrogen Bonds are Shown as Dotted Lines)

guanine (G) with cytosine (C) (Figure 1.4). It is the different combinations of these four bases along the backbone that encode different genetic information.

In general, DNA is divided into three categories, right-handed A- and B- forms, and left-handed Z-form (Figure1.5) (5). The Waston-Crick model structure is referred as the B-form DNA, which is the most stable form under physiological conditions and therefore it was chosen as our study model.

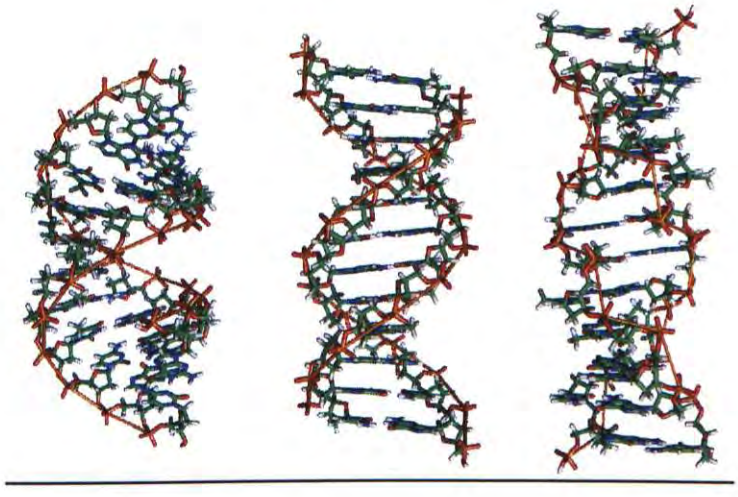


Figure 1.5 Structures of A-, B- and Z- form DNA (From Left to Right, Respectively)

To describe DNA structures conveniently and efficiently, some geometric parameters were introduced. Among them two kinds of parameters, the torsion angles and the helical parameters, are most important. The torsion angles describe the conformation of sugar-phosphate backbone. All of them, namely α, β, γ, δ, ε, ξ and υ_0, υ_1, υ_2, υ_3, υ_4, as

well as the orientation of the base χ, are defined as shown in Figure 1.6. The conformation of sugar ring, however, is usually described by another term, the pseudorotation phase angle P [1-1], which is given rise to the endocyclic torsion angles of sugar. Values of phase angles from 0° to 360° are given in multiples of 36° and hence ten conformations are included (Figure 1.7).

$$\tan P = \frac{(\nu_4 + \nu_1) - (\nu_3 + \nu_0)}{2\nu_2(\sin 36^{\circ} + \sin 72^{\circ})} \qquad [1\text{-}1]$$

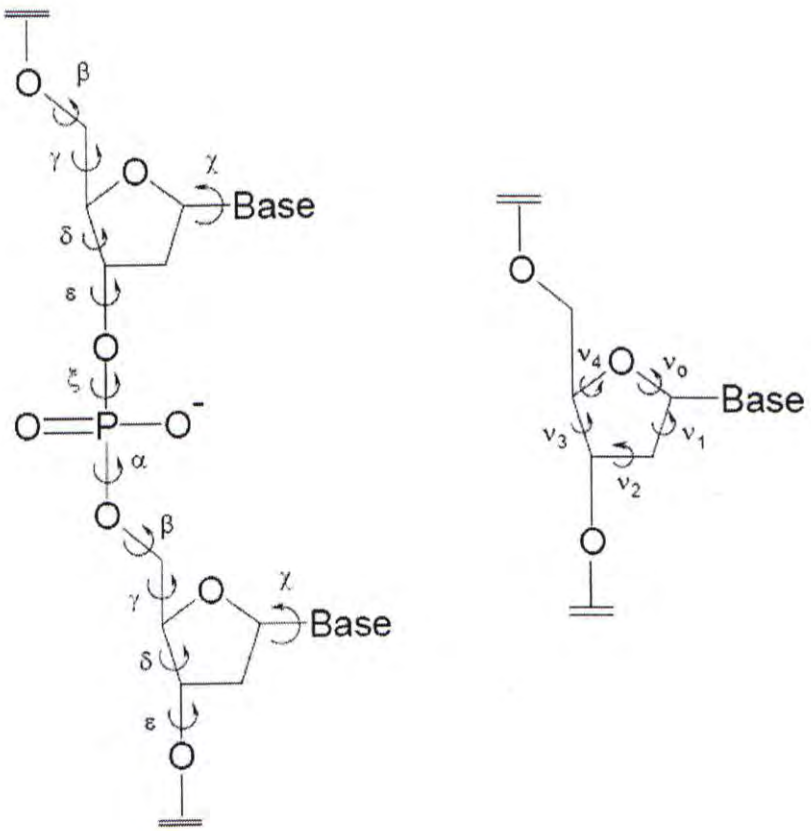


Figure 1.6 Definitions of the Torsion Angles of DNA Backbone and Sugar Ring

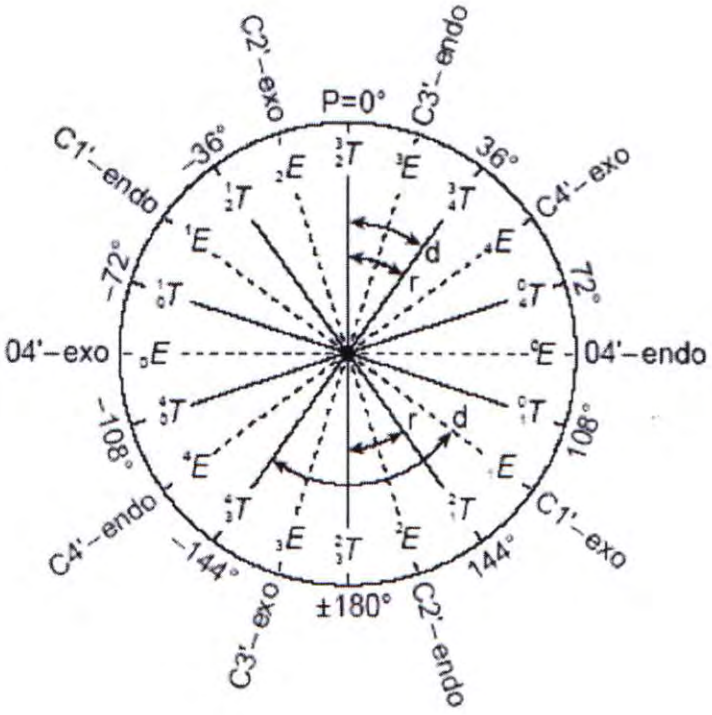


Figure 1.7 Definition of the Pseudorotation Angle (7)

The other most important kind of parameters, the helical parameters, are utilized to describe the geometry of nucleic acids.Totally 16 defined parameters are included according to the EMBO workshop (8). They are divided into two categories according to their motion modes, namely rotation mode and translation mode. Rotation mode includes the inclination (η), tip (θ), buckle (κ), propeller twist (ω), opening (σ), tilt (τ), roll(ρ) and twist (Ω). Translation mode includes x displacement (dx),y displacement (dy), shear (Sx), stretch (Sy), stagger (Sz), shift (Dx), slide (Dy), and rise (Dz) (Figure 1.8).

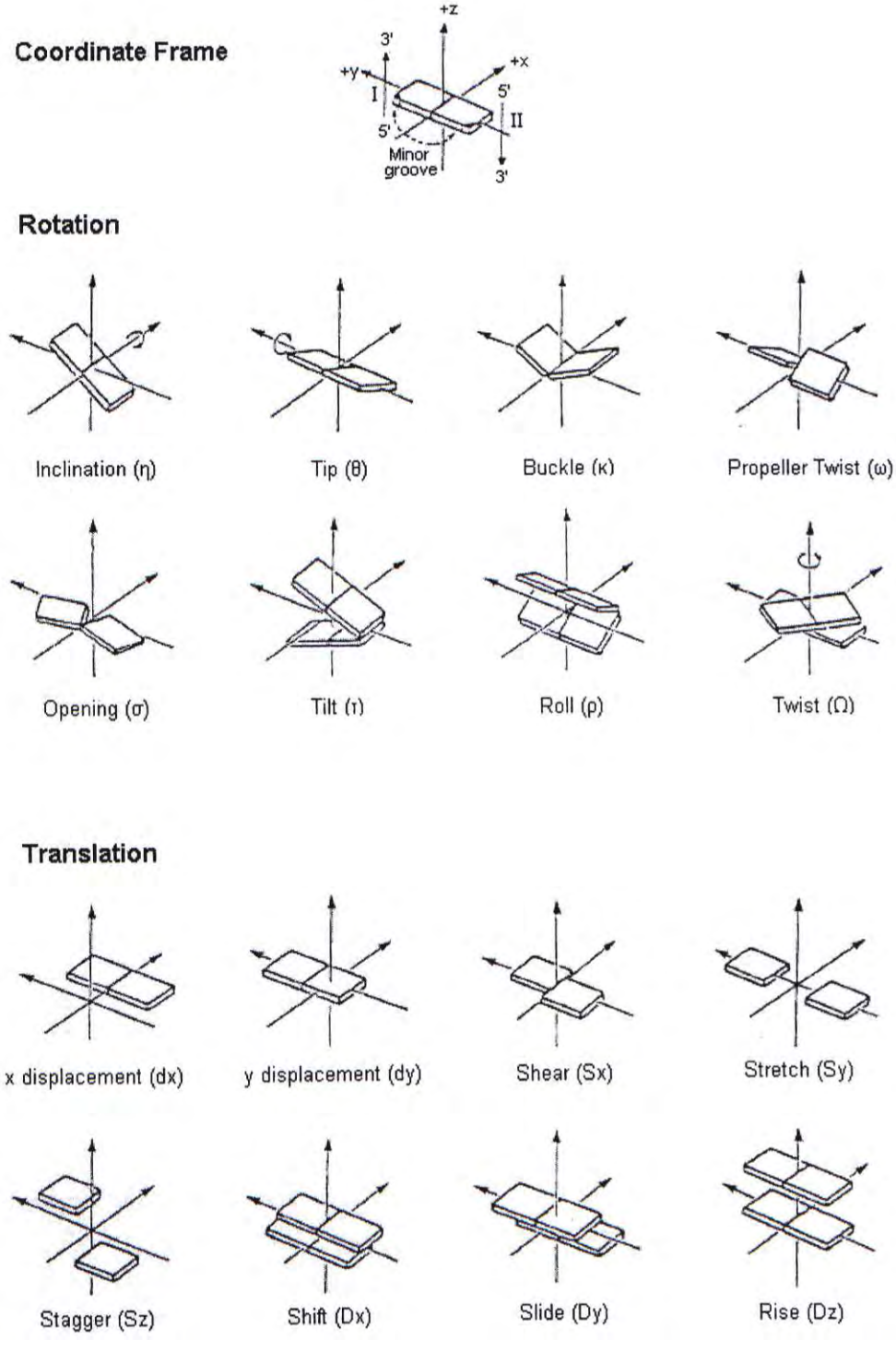


Figure 1.8 Definitions of the Helical Parameters

1.3 DNA Studies

DNA is so important that study on DNA is always of scientists' interest since the day on which it was discovered. An enormous body of experimental and theoretical data had been reported over the years (6, 9, 10). Usually the experimental data comes from X-Ray and NMR trials, and after obtaining the structure coordinates from experiments, theoretical investigation were carried out.

It's revealed that guanine's ionization potential (IP) is the lowest among the four bases (Table 1.1). Therefore guanine (G) shows the most electron donating property and is most easily oxidized. In the presence of oxidants or under UV irradiation, G is readily oxidized and ultimately forms 8-oxoguanine (or dihydro-8-oxoguanine) (11-13), which leads to mutation during replication (14). In addition, since G-rich sequence is more electronegative, as demonstrated by the lower IP values of consecutive G (i.e. GG, GGG, GGGG) comparing with that of single G (Table 1.1), the positive charge of G radical cation ($G^{+\cdot}$) can migrate to the G-rich position in the sequence through charge transfer (Figure 1.9). And consequently, the mutational hot spot changes.

For a G-rich sequence, the 5'-G is more electronegative than the 3'-G, which is demonstrated by the fact that the Highest Occupied Molecular Orbital (HOMO) of stacked contiguous G is largely localized on 5'-G (12, 13, 16). However, in some cases, such as

TGGGT, MO analysis indicates that the center G is the most reactive (17). It's suggested that the reactivity of DNA is a property of sequence dependence.

Table 1.1 Ionization Potential of N-Methylated Nucleobase Monomers and Stacked Contiguous Guanines (eV) (15)

Base	G	A	C	T	GG	GGG	GGGG
IP (eV)	7.75	8.24	8.87	9.14	7.28 (0.47)*	7.07 (0.68)*	6.98 (0.77)*

* The value in parentheses is the difference from the IP of single guanine (G)

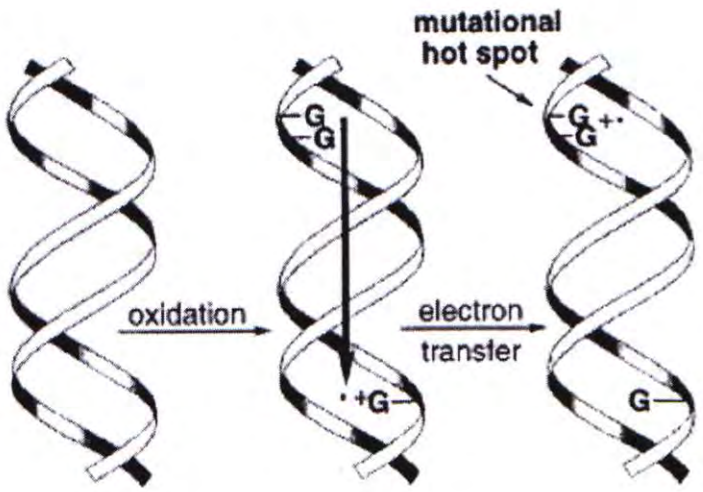


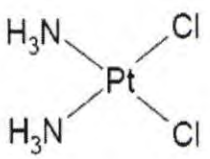
Figure 1.9 Oxidation of G and Charge Transfer to GG

G-rich sequences and oxidative damage are very important in biological systems, whereby the reactivity and the sequence pattern are highly related. G, especially, plays a key role in DNA damage. It's suggested that drug design to treat genetic diseases can be targeted towards the reaction with DNA bases. Further, understanding the structural and

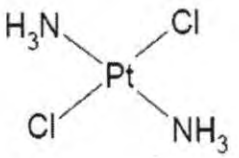
electronic properties of DNA is expected to assist in the recognition of the biological targets (18).

1.4 Cisplatin Studies

Cisplatin (Figure 1.10a) is a widely used anticancer drug that has been particularly successful in treating small cell lung, ovarian, testicular, head, and neck tumors (19). Since its anticancer activity was discovered by Rosenberg et al. in 1969 (20), Pt complexes have received much attention for their antitumor effect. Nowadays, although other anticancer drugs eventually take the place of cisplatin for their relatively low nephrotoxicity and drug-resistant, it is still used as a benchmark for new drugs (21). Due to this, it was chosen as a representative for chemotherapy anticancer drugs in our study.



(a)



(b)

Figure 1.10 (a) Cisplatin and (b) Transplatin

During the last 40 years, much progress has been made in understanding cisplatin's mode of action. Many details of the mechanism leading to antitumor activity have been well established (22-24). Now it's widely accepted that cisplatin's antitumor activity comes

from its binding to DNA and thus leading to an overall curvature of the DNA double helix. The distorted DNA then binds to the high mobility group (HMG) domains of proteins, impeding replication and cell repair processes, leading eventually to cell death (25).

In view of the case of DNA binding, there have been many studies of the reaction between cisplatin and DNA bases. Now it's certain that cisplatin firstly binds to the N7 site of purines, and prefers guanine over adenine. Platination to two purines gives the DNA-adducts with intrastrand and interstrand crosslinks. Experimentally, it's known that the bridged structure consists of 65% 1,2-d(GpG) intrastrand, 25% 1,2-d(ApG) intrastrand, and the rest other bridged structures like 1,3-d(GpXpG) intrastrand, the d(GpC)d(GpC) interstrand etc. Interestingly, the adduct of 1,2-d(GpA) intrastrand is hard to find (26, 27). Binding of cisplatin had been shown to take place on the A of the 5' side of d(ApGpA) (28). It's suggested that the failure to yield cisplatin-d(GpA) bridged product is due to the slow monofunctional binding of cis-$[PtCl(NH_3)_2(H_2O)]^+$, whereby the binding is the slowest when compared to that of GG or AG (29).

The anticancer activity of cisplatin as mentioned above comes from its binding with purines of DNA and leading to a structural distortion of the DNA and finally causing cell death. In the past years, scientists did a lot of work to theoretically investigate the mechanism of the binding of DNA with cisplatin. Maybe due to the huge molecular weight

of whole DNA and thus the huge computational cost, however, most of the theoretical studies focused on the interaction between cisplatin and single base, single base-pair or just a few base-pairs, and most of the work even excluded the sugar-phosphate backbone out of their scope. To mimic a more realistic situation, it's very necessary to construct a case of longer DNA sequences, including sugar-phosphate backbone, solvent effect, electrostatic effect etc.

1.5 Scope of the Thesis

Fu did a meaningful job on theoretically investigating the structures and the electronic properties of DNA, and its ligating properties with cisplatin, as well as the factors controlling the formation of the binding products (18). In her study, a series of short segment DNA, namely, 5'-GX-3', 5'-AX-3', 5'-XGG-3', 5'-GGX-3', 5'-GXG-3', 5'-XAG-3', 5'-AXA-3', 5'-XGGX-3', were optimized and studied using *ab initio* method in gas phase. As a conclusion, she showed that HOMO localizes on G for any G-containing DNA sequence. In addition, a set of empirical selection rules was suggested for identifying the locations of HOMO in DNA sequences. Further, she found the chelation of cisplatin is selective and sequence-dependent (18).

However, all of her work were based on the optimized structures of DNA models at the RHF/STO-3G level of theory in gas phase. It suffers from some disadvantages. Firstly,

the realistic biochemical environment of DNA is more complicated than in gas phase. Not only longer DNA sequences is needed, but also many effects such as solvent effect, electrostatic effect etc. should be considered. Thus the results obtained in gas phase may be not accurate enough to draw practical conclusions. Secondly, the basis set STO-3G used is the simplest basis set. It is necessary to use a more advanced basis set when dealing with the situation with various effects (such as polar effect etc.). Moreover, *ab initio* method (such as RHF) is computationally demanding and time consuming (i.e. very expensive) and thus not suitable for the computation of realistic systems. Subsequently, alternative time saving strategy needs to be developed.

Nowadays, a number of computational methods are available whereby some of these methods have been modified to handle large system (30-35). Unfortunately, some of the DFT based methods, e.g. plane wave DFT (31) cannot be used in the MO analysis as they cannot be used together with NBO methods. For the structure optimization, although new DFT (32) and semi-empirical (34) methods are available, benchmark testing (accuracy vs. time) as to the suitability to the large systems has not been carried out. In this project, a hybrid Molecular Dynamics + Quantum Mechanics (MD+QM) method (25, 36-37) was utilized. DNA models, as well as the solvent molecules (water) and the counter ions (Na^+, to neutralize the system), were firstly equilibrated (optimized) using molecular dynamics

(MD) method; after then MO analysis was carried out with quantum mechanics (QM) method and relative information was then derived or deduced.

The ultimate goal of this project is to re-examine the set of rules proposed by Fu (18) which uses an established chemical basis to predict the formation of Pt-DNA lesion. In Chapter 2, computational details were given and the applicability of the MD+QM method was discussed. In Chapter 3, the HOMO and HOMO-1 (the next Highest Occupied Molecular Orbital) location obtained by the proposed computation strategy, of the series of short segment DNA adopted in Fu's thesis (18), was re-examined and analysis was then followed. Moreover, the analysis would be extended to systems with 5 to 8 base pairs, especially sequences with pure G such as GGGGG, GGGGGG, GGGGGGG, GGGGGGGG. This can particularly shed light on the electronic preference of G1-G2 binding versus of G1-G3 binding. The possible reaction between cisplatin and those DNA models would be also discussed from structural point of view in this chapter. Then in Chapter 4, the obtained results would be summarized and some conclusion would be drawn for the investigation work.

CHAPTER TWO

METHODOLOGY AND COMPUTATION

2.1 Introduction

Nowadays, much computer-based software had been developed in computational chemistry. However, looking for an appropriate approximation is still a preoccupation to quantum mechanical calculation. As stated in Chapter 1, a hybrid Molecular Dynamics + Quantum Mechanics (MD+QM) method was utilized in this project. Details of the method are described as follows.

2.2 Molecular Dynamics Simulation

Molecular Dynamics (MD) is a simulation of the time-dependent behavior of a molecular system. It uses molecular mechanics (MM) method to compute the energy of a system. The information about possible energy levels and conformations could be obtained by this method (38, 39).

AMBER software was used to run MD of the DNA models in this work. Generally, AMBER refers to two things: a set of molecular mechanical force field for the simulation of bimolecular and a package of molecular simulation program which includes source code and demos. Here the latter meaning is referred. The version of the AMBER program used in this work is version 9, which was released by University of California,

San Francisco (UCSF) subject to a licensing agreement (40). The computational procedures as well as the details are listed as follows.

The initial structures of the DNA models in B-form double helix were constructed with the *nucgen* script in AMBER package. Noting that DNA models are anions due to the ionized phosphate groups, *tleap* script was subsequently adopted to neutralize the systems using sodium counter ions. A truncated octahedral box of water, TIP3PBOX, around the DNA models with the radius of 8 Å, was then created. The topological files and the coordination files were saved for the next minimization process.

The minimization process includes two stages. During the first stage DNA molecular, excluding the sodium ions and the water molecules, was fixed using a 500 $kcal \cdot mol^{-1} \cdot Å^{-2}$ force constant. Particle Mesh Ewald (PME) method in conjunction with constant volume periodic boundaries was used to calculate long range electrostatic interaction, with a cutoff of 10 Å. 500 steps of steepest descent minimization followed by 500 steps of conjugate gradient minimization were then carried out. After completion the first stage ends up and the second stage shall start. During the second stage, the entire system was minimized. No component was fixed this time and just let the whole system relax. 1000 steps of steepest descent minimization followed by 1500 steps of conjugate

gradient minimization were then carried out. After completion, the whole minimization process was accomplished and the next molecular dynamics process was followed.

The molecular dynamics (MD) process also includes two stages. During the first stage the system was heated from 0 K to 300 K with weak restraints on the solute. Random initial velocities from a Boltzmann distribution were then generated and coordinates from the coordination files were read. PME method in conjunction with constant volume periodic boundaries was used. SHAKE keyword was turned on to constrain bonds involving hydrogen. The langevin dynamics was used to control the temperature with a collision frequency of 1.0 ps^{-1}. 10,000 molecular dynamics steps with 2 fs per step were carried out, to give a sub-total simulation time of 20 ps.

After successfully heating the system at constant volume with weak restraints on the DNA, it is necessary to switch to constant pressure (1 atm, isotopic position scaling was used to maintain the pressure, and 2 ps relaxation time was needed) so that the density of water can relax. Meanwhile, since it is already at 300 K, the restraints on the DNA can be safely removed. 100 ps was allotted for the second stage, namely equilibration stage to give the whole system plenty of time to relax.

MD simulation was accomplished after completion of all above procedures. Before going on, however, it is essential to check whether equilibrium has been successfully achieved or not. A number of system properties, namely volume, density, temperature, pressure, total energy, kinetic energy, potential energy and root mean square deviation (Rmsd), were selected to be examined. Figure 2.1 shows the results of a selected DNA model (GGG) from the simulations.

Notice how the volume of the system initially decreases as the water box relaxes and reaches a stable value (Figure 2.1). Smooth transitions in the plot followed by the slight oscillations around a mean value suggest the equilibration process was successful. So is the situation of the density of the system. It's a mirror of the volume. The system was lastly equilibrated at a density of about 1.02 g/cm^3. It is reasonable. Note that the density of pure liquid water at 300 K is approximately 1.00 g/cm^3. Adding a 3-mer DNA and associated charges has increased the density by around 2%.

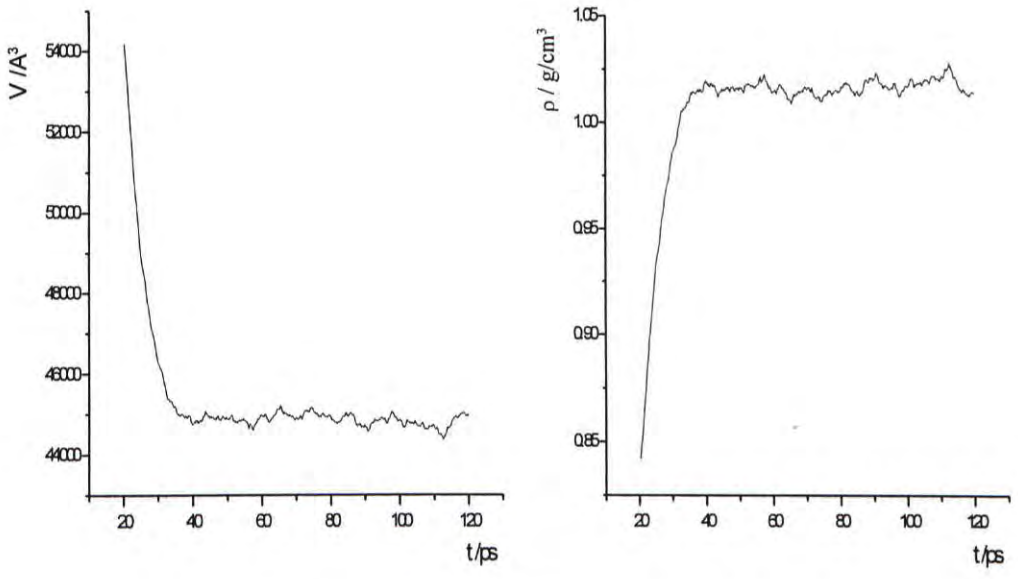


Figure 2.1 Volume (Left) and Density (Right) against Time Plots throughout the Simulation

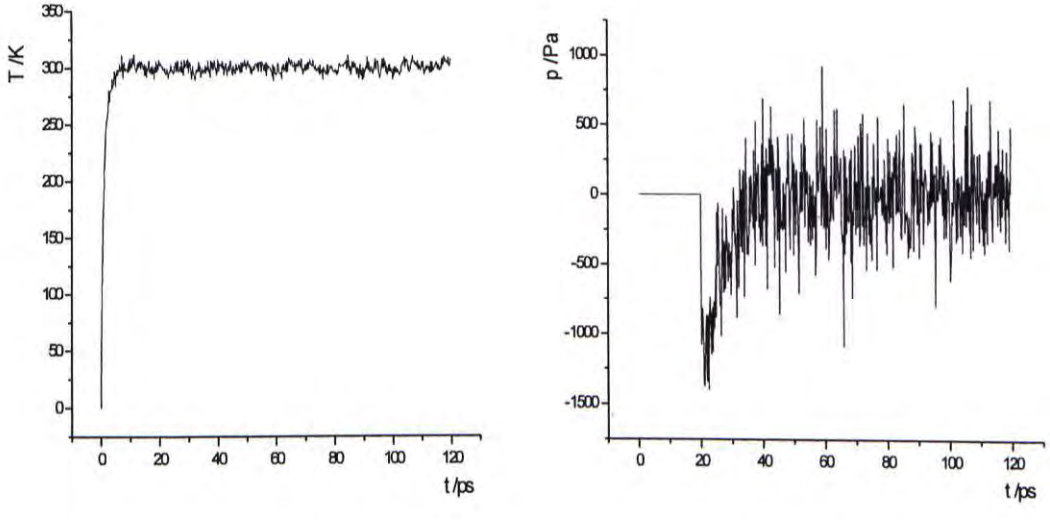


Figure 2.2 Temperature (Left) and Pressure (Right) against Time Plots throughout the Simulation

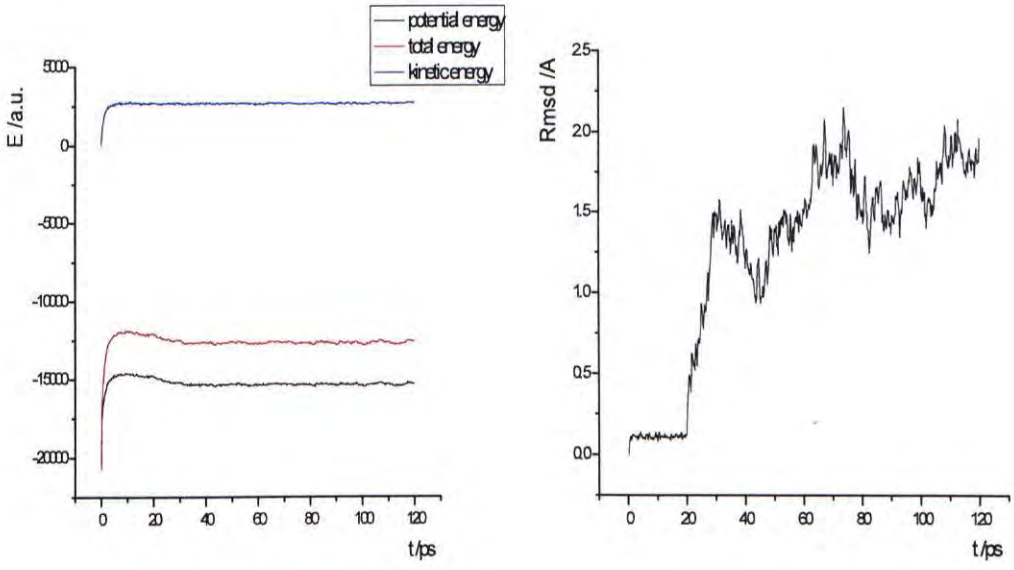


Figure 2.3 Energy (Left) and Rmsd (Right) against Time Plots throughout the Simulation

For the temperature of the system (Figure 2.2), it increased from 0 K to 300 K during the first 5 ps and then remained almost constant. It indicates that the use of Langevin dynamics for temperature regulation was successful. The situation is a little different for the pressure of the system (Figure 2.2). During the first 20 ps the pressure remains zero. This is expected since a constant volume simulation was running in which the pressure wasn't evaluated. At the time point of 20 ps it was switched to constant pressure, allowing the water box to relax. The pressure immediately dropped sharply and became negative. From 20 ps to 40 ps, the pressure fluctuated wildly. But after 50 ps, the

pressure eventually approached around 1atm. This validates that the equilibration was achieved.

In the energy plot (Figure 2.3), the red line, the blue line, and the black line represent the kinetic energy (KE), the potential energy (PE) and the total energy (E), respectively. It's noted that all of the energies increased during the first few picoseconds, corresponding to the heating event from 0 K to 300 K. The kinetic energy then remained constant for the rest time. It implied that the temperature thermostat program, which directly decides the kinetic energy, worked correctly. The potential energy, and consequently the total energy, initially increased and then stayed constant during the constant volume stage (0 to 20 ps). As it was switched to constant pressure, the PE and the E of the system decreased a little (20 ps to 50 ps) and then leveled off for the rest time (50 ps to 120 ps). In the Rmsd plot (Figure 2.3), it was seen that the Rmsd of the DNA backbone atoms remained low due to the restraints on DNA for the first 20 ps. After removing the restraints the Rmsd shot up as the DNA relaxed within the solvent (20 ps to 50 ps). After that the Rmsd is fairly stable with no wild oscillations (50 ps to 120 ps).

All the results suggested the simulation was successful. One can also take a look at the trajectory file to observe how the dynamics of the DNA molecules changed

throughout the simulation. This may also offer some additional clues for the quality of the simulation.

After completing all of this, it's ready to proceed to the next process, the quantum mechanics calculation.

2.3 Quantum Mechanics Calculation

Although molecular dynamics simulation is fast and time-saving, it suffers from several limitations such as requiring extensive parameters, inaccurate energy value and so on. It is still necessary to use it with the combination of the quantum mechanics method.

Although some research groups had successfully applied SCF, MP2, DFT methods to the study of the DNA base pairs (21, 24, 25, 41-44), while other groups used Fragment Molecular Orbital (FMO) scheme to treat large scale molecules (45-47), Hatree-Fock method (HF), however, remains the primary method of the choice. This is on the one hand due to its lower computational cost (compared with SCF, MP2, DFT methods); on the other hand making the comparison of the data with that obtained using pure HF method reasonable.

The applicability of HF method for DNA models had been discussed by Fu (18) and won't be presented here. Except for the computational method, the choice of basis function is also important. Generally speaking, the larger the basis set, the more accurate the results, but meanwhile the more expensive the computation. Due to the limitation of current computational resources, 3-21G basis set is ultimately chosen. Considering the polarization effect, 6-31G* basis set is adopted for the phosphorus atom (P) separately.

Since there are too many atoms (up to some thousands) in the system, it's almost impossible to carry out an *ab initio* calculation. Therefore water molecules were tripped from the system after molecular dynamics simulation process. To get the optimized structure, the coordinates from 50 ps to 120 ps of the molecular dynamics was averaged since there is no wild fluctuation for Rmsd value (Figure 2.3) during this period. To neutralize the system, sodium ions (Na^+) were kept. For simplification, only singlet states of DNA models were considered.

A single point energy (SPE) calculation was first done. Then, to elucidate the electronic properties of the DNA, the HOMO and HOMO-1 were located. All of the quantum mechanics calculation was performed using Guassian03 program (48) at

RHF/GEN (3-21G for C, H, O, N, Na; 6-31G* for P) level of theory using PC clusters and workstations.

2.4 Verification of Methodology

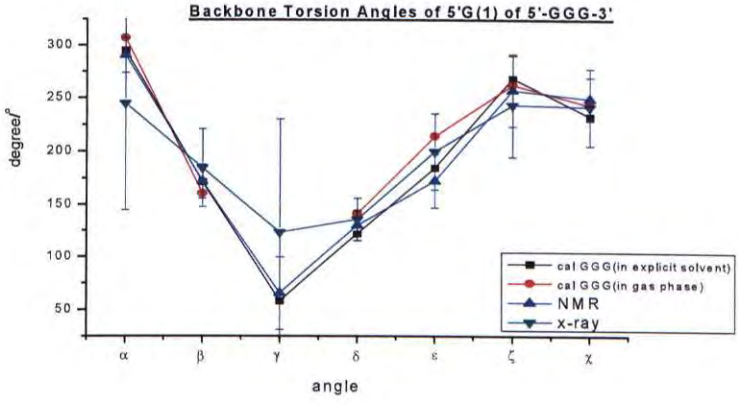
To verify the method utilized, it is necessary to compare the results with the experimental data. Fu suggested three parameters, namely the backbone torsion angle, N7-N7 distance and the location of HOMO for the verification (18). Here her suggestion was adopted and the comparison details for each parameter are listed below.

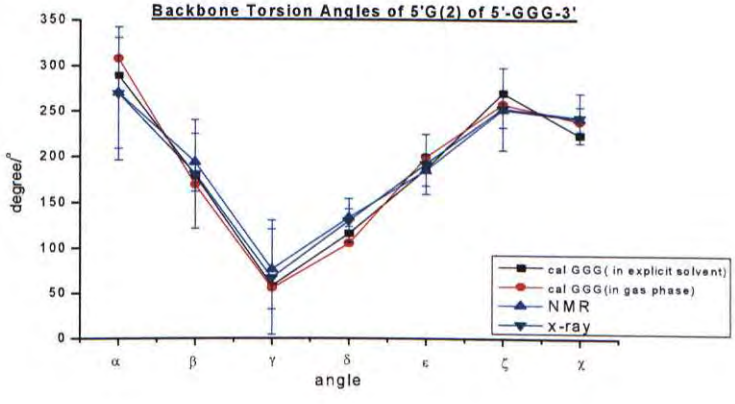
2.4.1 *Backbone Torsion Angles*

The sugar-phosphate backbone torsion angles of the optimized three base-pair structure 5'-GGG-3' and four base-pair structure 5'-GGGG-3', were compared with the experimental results obtained from the Nucleic Acids Database (NDB) (49). Structures of 47 sequences determined using X-ray (50) and 12 sequences determined using NMR (51) containing the GGG fragment were extracted for comparison (Appendices I and II). Moreover, to have a comparison between the results calculated in solvent and those calculated in gas phase, the points obtained by Fu (18) were also plotted.

For the case of 5'-GGG-3', it was found that all the torsion angles match the experimental data well (Figure 2.4). So is the case of 5'-GGGG-3' (Figure 2.6). Since one phosphate group per strand was omitted when DNA models were constructed by the AMBER program or SYBYL 6.2 program in Fu's study (18), some parameters, i.e. α, β, ε, ξ of the 3'-end base such as the third base of 5'-GGG-3' were not available. To partly validate whether the parameters of the third base match the experimental data or not, a comparison between the parameters of computed 5'-GGGG-3' and those of experimental 5'-GGG-3' was carried out (Figure 2.5). And the result shows that wherever the base is, the data between the corresponding bases is compatible.

Moreover, it was found that there is more or less difference between the data calculated in solvent and that in gas phase. In most cases the data obtained in solvent is more close to the experimental value than that obtained in gas phase. Even though for the terminal base, the data is more "reasonable" in solvent. This result is inspiring. Not only does it implies an appropriate system (DNA model + counter ion + solvent) was chosen and a successful simulation was carried out, but also it primarily validates the high application of the hybrid MD+QM method to the system.





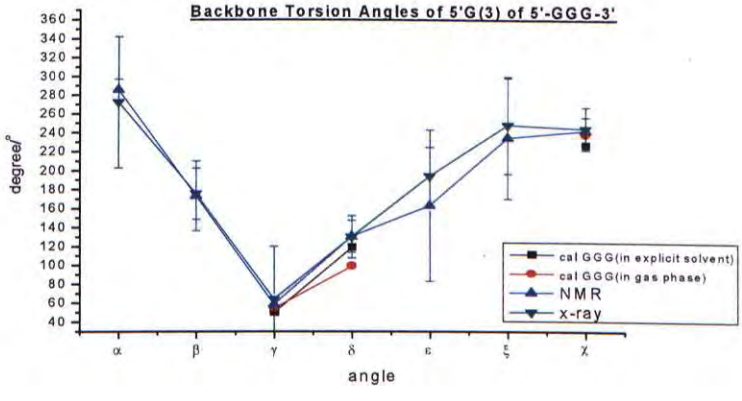


Figure 2.4 Comparisons of Backbone Torsion Angles of Computed and Experimental 5'-GGG-3'

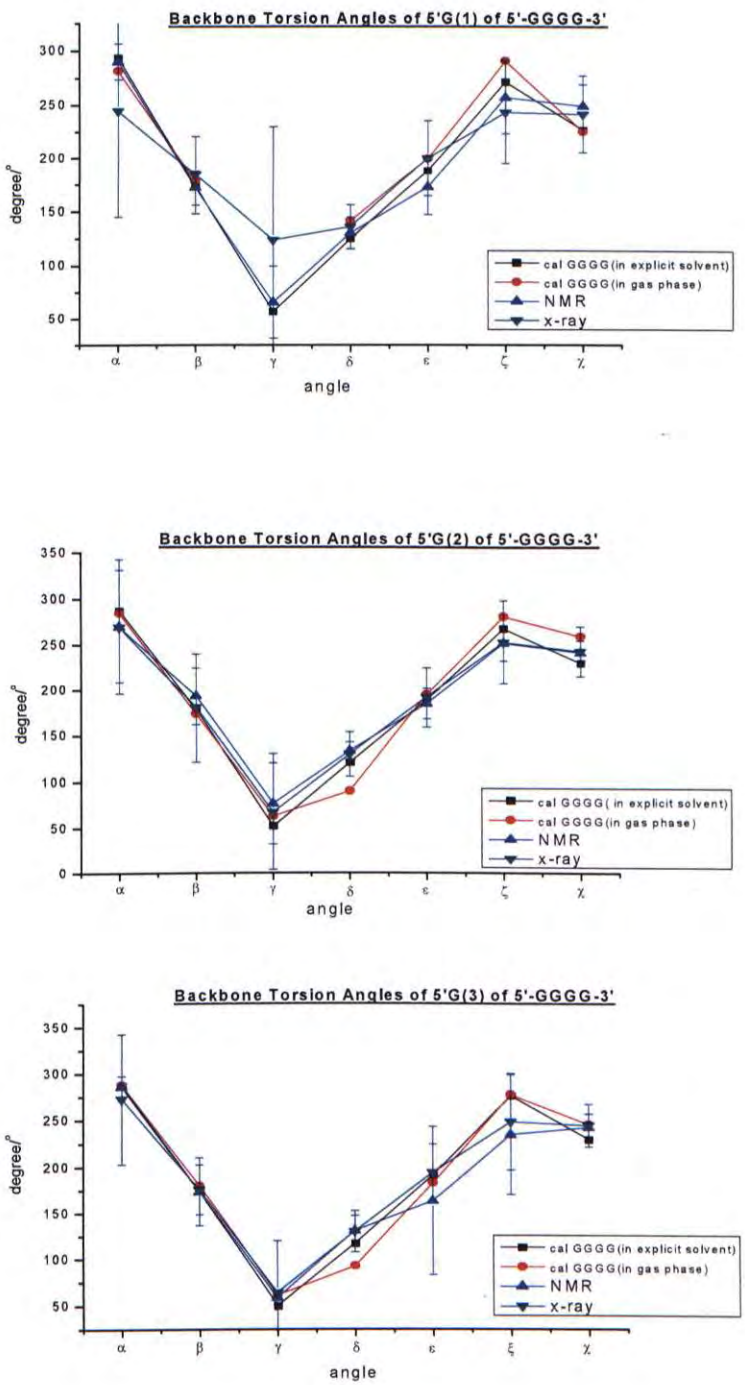
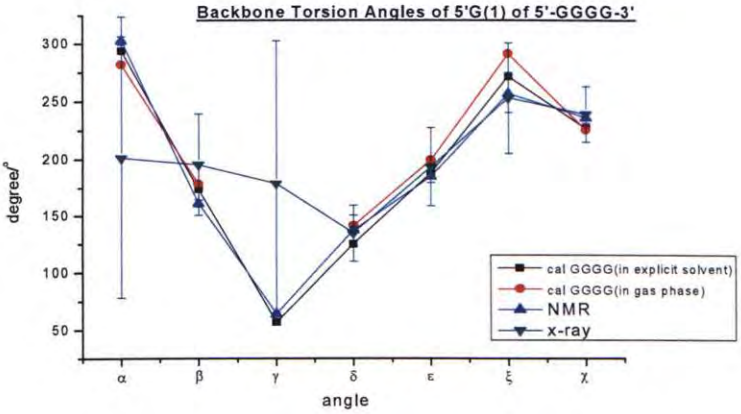
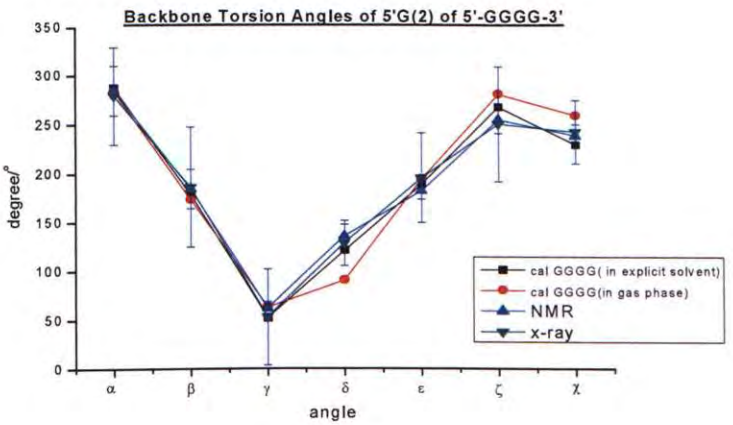
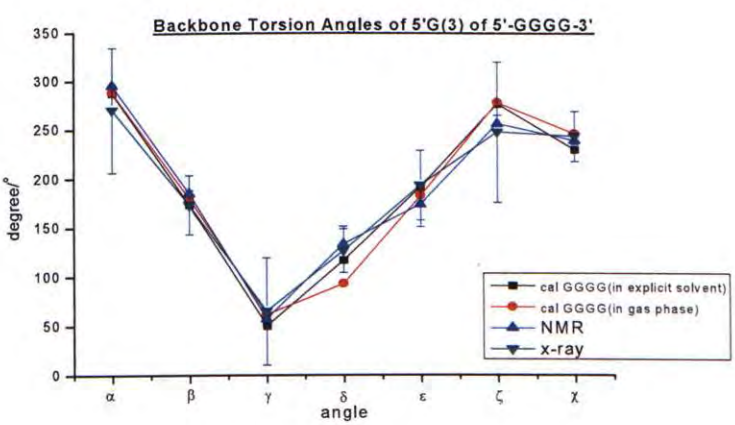


Figure 2.5 Comparisons of Backbone Torsion Angles of Computed 5'-GGGG-3' and Experimental 5'-GGG-3'


Backbone Torsion Angles of 5'G(1) of 5'-GGGG-3'
degree/°
300
250
200
150
100
50
α
β
γ
δ
ε
ξ
χ
angle
cal GGGG(in explicit solvent)
cal GGGG(in gas phase)
NMR
x-ray


Backbone Torsion Angles of 5'G(2) of 5'-GGGG-3'
degree/°
350
300
250
200
150
100
50
0
α
β
γ
δ
ε
ζ
χ
angle
cal GGGG(in explicit solvent)
cal GGGG(in gas phase)
NMR
x-ray


Backbone Torsion Angles of 5'G(3) of 5'-GGGG-3'
degree/°
350
300
250
200
150
100
50
0
α
β
γ
δ
ε
ζ
χ
angle
cal GGGG(in explicit solvent)
cal GGGG(in gas phase)
NMR
x-ray

(cont'd)

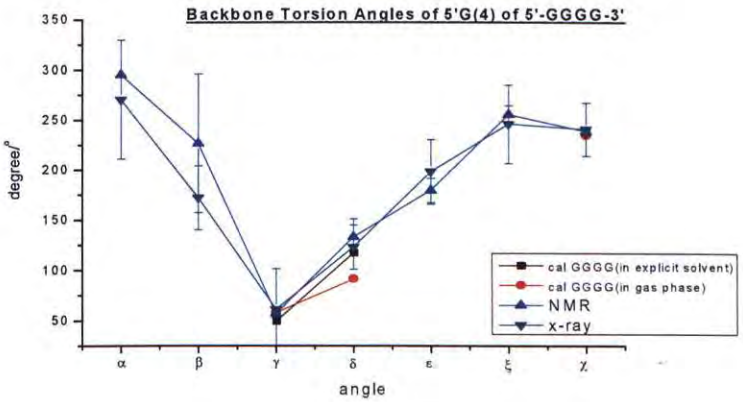


Figure 2.6 Comparisons of Backbone Torsion Angles of Computed and Experimental 5'-GGGG-3'

2.4.2 *N7-N7 Distance*

As a reference, N7-N7 distance of two adjacent guanines was measured for the optimized structures GG, GGG, GGGG, GGGGG, GGGGGG, GGGGGGG, GGGGGGGG (Table 2.1). It's found that the average distance between the neighboring guanine varies from 3.91 Å to 4.14 Å, while the maximum distance is 4.7234 Å and the minimum 3.5942 Å. The data is also very close to the experimental value. According to Astbury (3), the inter base-pair distance of the first X-ray photographs of fibrous DNA is 3.4 Å. Fu suggested an average distance of 4.23 Å from some X-ray and NMR sequences in her study (18). For the case of dodecamer d(CGCGAATTCGCG), the mean rise

distance is 3.33 Å (9) and for B-form DNA generally, the stacked distance is 3.38 Å. Although the calculated distance is more or less larger than the experimental one, in contrast with the data calculated in gas phase the average value of which varies from 4.95 Å to 6.35 Å (Table 2.1), great progress was achieved.

Table 2.1 N7-N7 Distance of the Optimized Structures of GG, GGG, GGGG, GGGGG, GGGGGG, GGGGGGG, GGGGGGGG

	GG[a]	GGG	GGGG	G5[e]	G6	G7	G8
G1-G2[c]	4.0949(6.35)[b]	3.9329(5.54)		3.7035	4.3431	4.7040	4.1762
G2-G3	--	4.0538(5.55)	3.9813(4.16)	**4.7234**	3.9046	3.7437	4.6653
G3-G4	--	--	3.7596(5.29)	3.8397	4.4587	4.0353	4.1182
G4-G5	--	--	--	3.8208	3.7342	4.1477	3.9866
G5-G6	--	--	--	--	3.7994	4.3688	4.3271
G6-G7	--	--	--	--	--	**3.5942**	3.8310
G7-G8	--	--	--	--	--	--	3.8397
Average[d]	4.0949(6.35)	3.9934(5.55)	3.9133(4.95)	4.0218	4.0480	4.0989	4.1349

a The sequence direction is from 5' terminal to 3' terminal

b The value in bracket showing the results calculated in gas phase with pure quantum mechanics method

c Base pair step, starting from the 5' terminal

d Average value of G1-G2, G2-G3, G3-G4, G4-G5, G5-G6, G6-G7, G7-G8, if applicable

e G5 represents 5 consecutive G, so and so forth

Furthermore, the average distance doesn't become smaller as expected but disordered as the sequences become longer (Table 2.1). This result implies that the hybrid method doesn't depend on the length of the sequence. But for the calculation in gas phase

with pure quantum mechanics method, this is not the case (18). The N7-N7 distance becomes smaller as the DNA sequence becomes longer. Based on this, Fu proposed a scaling factor deducing from 2, 3 and 4 base pairs to predict the distance value of sequences with other lengths. She also concluded that DNA models with a minimum of six base pairs are necessary and models with twelve base pairs are the best representation of B-DNA. But during the molecular dynamics process, DNA models accompanying with the counter ions and the solvent molecules have enough time to relax. This "simulates" a more realistic situation and may be responsible for the disordered average N7-N7 distance. Also note that the DNA models calculated in gas phase was constructed using SYBYL 6.2 (18, 52), but the models calculated in solvent was built using AMBER 9 (40). This may also have some influence but the effect may be not that vigorous.

In summary, like the situation of backbone torsion angle, N7-N7 distance is more reasonable calculated in solvent with the hybrid MD+QM method than in gas phase with pure quantum mechanics method. In addition, the data in solvent has no obvious correlation with the length of the DNA sequence, which is different from the situation in gas phase. For 2 to 8 consecutive guanine models, the fluctuating range of the average

distance is just 0.23 Å. It implies no matter how long the DNA models are, the results always remain reasonable.

2.4.3 *Location of HOMO*

To further verify the applicability of the method, another comparison of locating the MO with RHF/STO-3G method (18) in gas phase was carried out. Figure 2.8 lists the HOMO comparison results of AG, GA, GG and AA models. The HOMOs from the two methods match very well. Fu once had a comparison of her obtained HOMOs with the experimental results and found they are compatible (18). She also had a calculation using a higher level of theory, namely UB3PW91/6-311G(d, p), and it was found that in spite of the difference in the shape of the HOMOs obtained by two computational methods, the major position of HOMOs is same. Since the HOMO results obtained by different methods are compatible, it once again indicates the hybrid MD+QM approach is practical and appropriate for illumination for the systems.

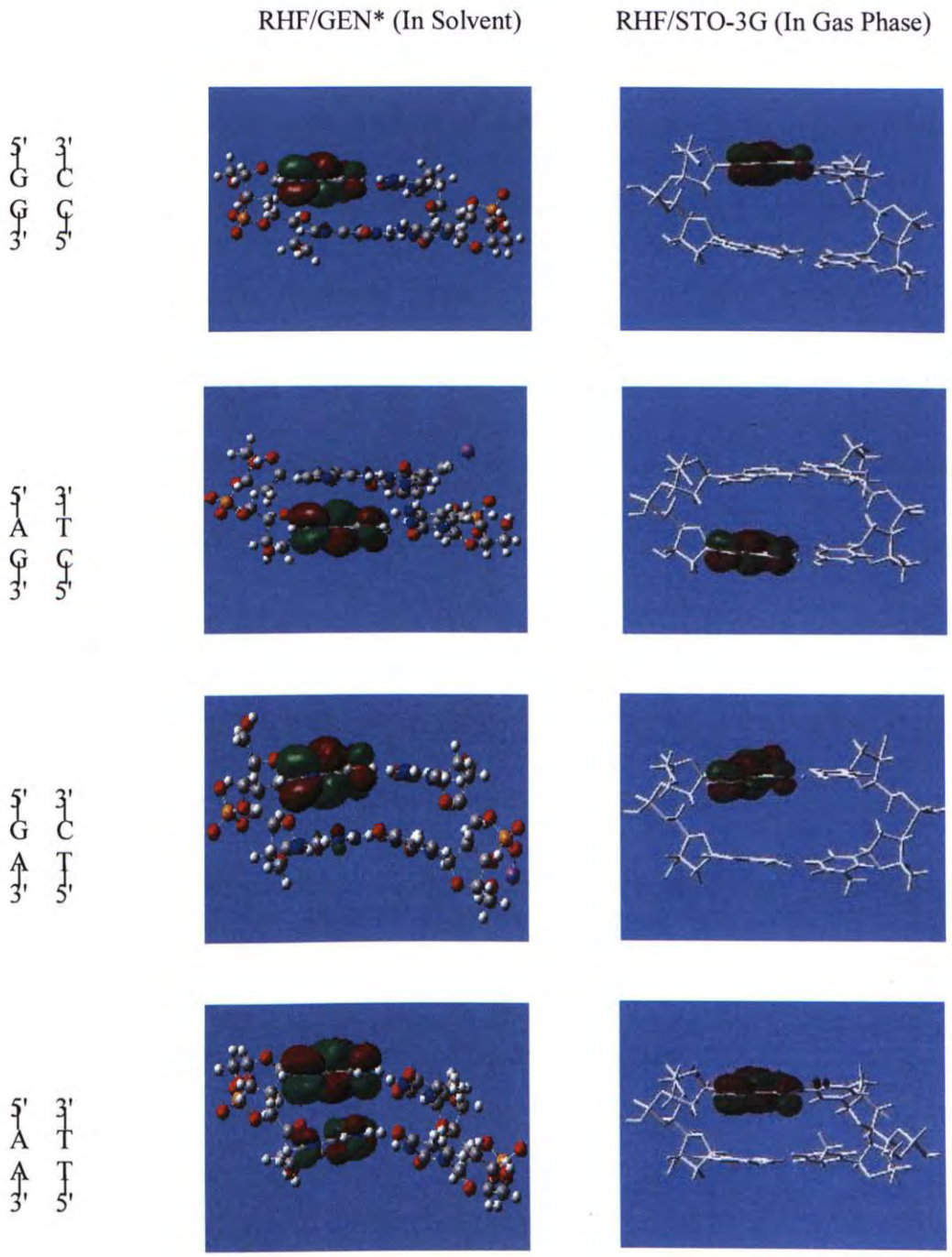


Figure 2.7 Comparisons of Locations of HOMO of 2 Base-Pair Models Computed by Hybrid Method in Solvent and Pure Quantum Mechanics Method in Gas Phase

* 3-21G basis set for C, H, O, N, Na; 6-31G* basis set for P

2.5 Summary

In this chapter, a hybrid MD+QM method was applied to the systems containing DNA models, counter ions and solvent molecules. This method is found to be really timesaving during the calculation. For an 8-mer DNA model, which is neutralized by 14 Na^+ and surrounded by about 2,000 water molecules, it costs only about 1.5 days for a 120 ps molecular dynamics simulation and about 0.5 days for a quantum mechanics calculation (water molecules were tripped from the system for *ab initio* calculation). Besides, by comparing the computed (in solvent and in gas phase) and experimental data of the backbone torsion angles, the reference distances, as well as the locations of HOMO, it's found that the results obtained by the hybrid method approaches the experimental data more closely than those obtained by pure quantum mechanics method in gas phase. It is concluded that this method can at least semi-quantitatively calculate the electronic structures of DNA and it is thus adopted for the further study.

CHAPTER THREE

UNDERSTANDING OF THE CISPLATIN-DNA CROSSLINKS

3.1 Introduction

Now it's well established that the anti-tumor activity of cisplatin resulted from binding to DNA bases at the N7 position of the imidazole ring of guanine (G) or adenine (A) (23) and that binding is preferentially at G (27, 53). In the past few years, much research work focused on the subtle chemical process of the interaction between cisplatin and DNA, such as charge transfer (54, 55), hydrogen bonding pattern (42, 56, 57), proton transfer (58), hole migration (59), π-stacking interactions (60) and so on. Less work, however, was carried out towards the interaction between cisplatin and whole DNA sequences. If the DNA sequence changes, the impact on its interaction with cisplatin is unknown. Fu did a precursory and meaningful job with respect to this in 2004 (18). In her research, a series of DNA models was calculated at the RHF/STO-3G level and based on that a set of rules was raised up to predict the binding site where cisplatin would connect. Although her work is promising, the results suffer from the systematic disadvantage due to the computational method. In last chapter, a hybrid method, namely MD+QM method, was demonstrated to be highly applicable to the DNA models. In this chapter, the hybrid

method was employed in determining (i) the potential binding products of cisplatin-DNA, and (ii) the structural investigation of the DNA models from a theoretical perspective.

3.2 MO Analysis

In the ligand chemistry theory, a ligand substitution reaction occurs due to the interaction between the Highest Occupied Molecular Orbital (HOMO) of the ligand and the Lowest Unoccupied Molecular Orbital (LUMO) of the substrate (61). In the case of cisplatin-DNA binding, it's thus expected that the electrons of HOMO (for monofunctional), or with the electrons of HOMO-1, i.e. the next Highest Occupied Molecular Orbital (for bifunctional) of DNA donate to the LUMO of cisplatin. As stated in Chapter 2, generally the HOMO of DNA locates on guanine bases. To further validate this and investigate the impact of sequence effect on the MO location, in this study, the HOMO and HOMO-1 of selected DNA sequences were located and a new rule was proposed to predict the binding product.

3.3 Potential Binding Products with the Ligand

The possible binding products can be predicted from the location of HOMO and HOMO-1 of DNA as stated above. Five kinds of crosslinks can be formed from the calculated results (Figure 3.1).

(a) 1,2-d(GpG) Intrastrand Crosslink

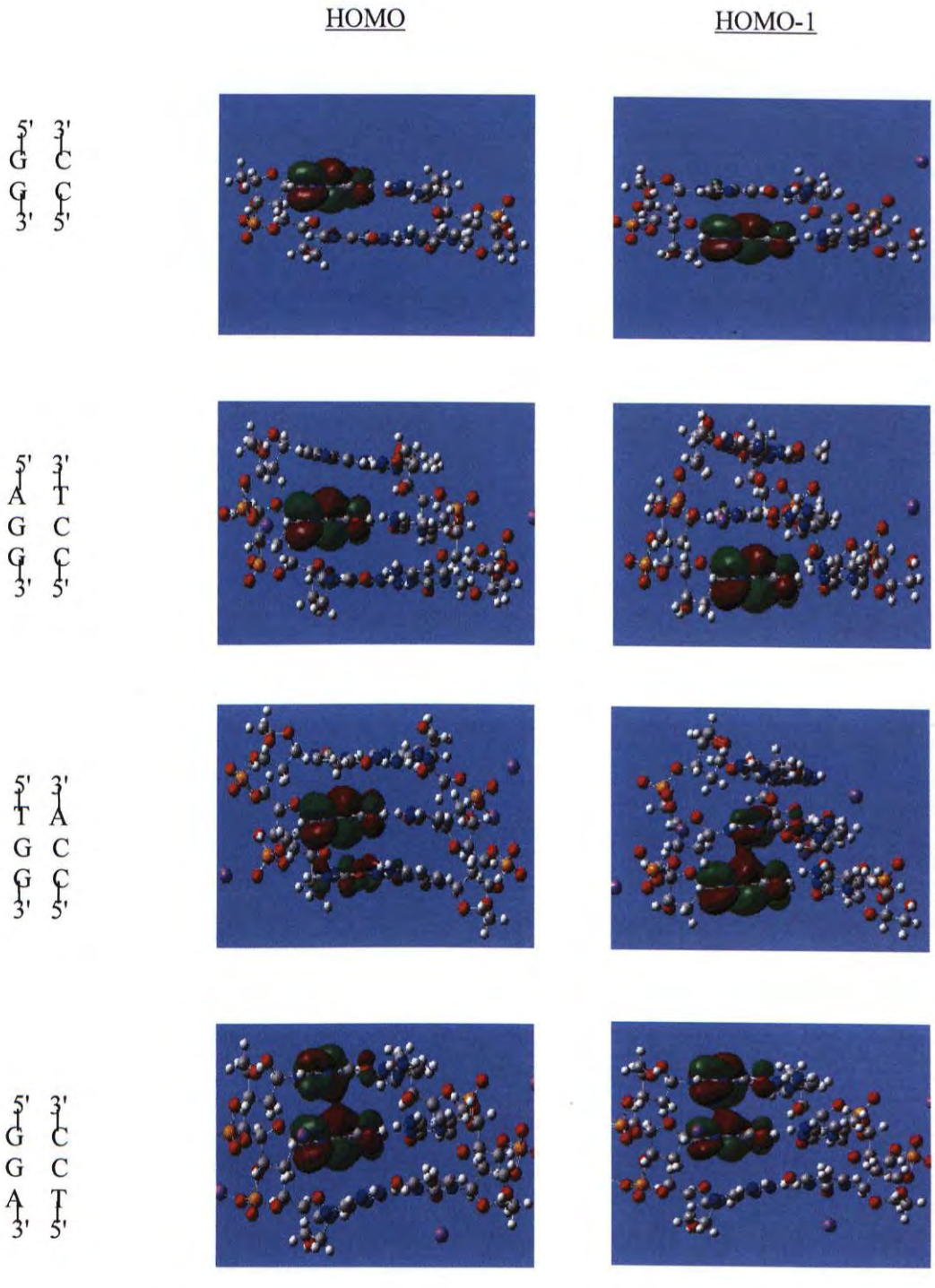


Figure 3.1 The Location of HOMO and HOMO-1 of the Base-pair Models, and the possible Crosslink to Cisplatin

(a) 1,2-d(GpG) Intrastrand Crosslink (cont'd)

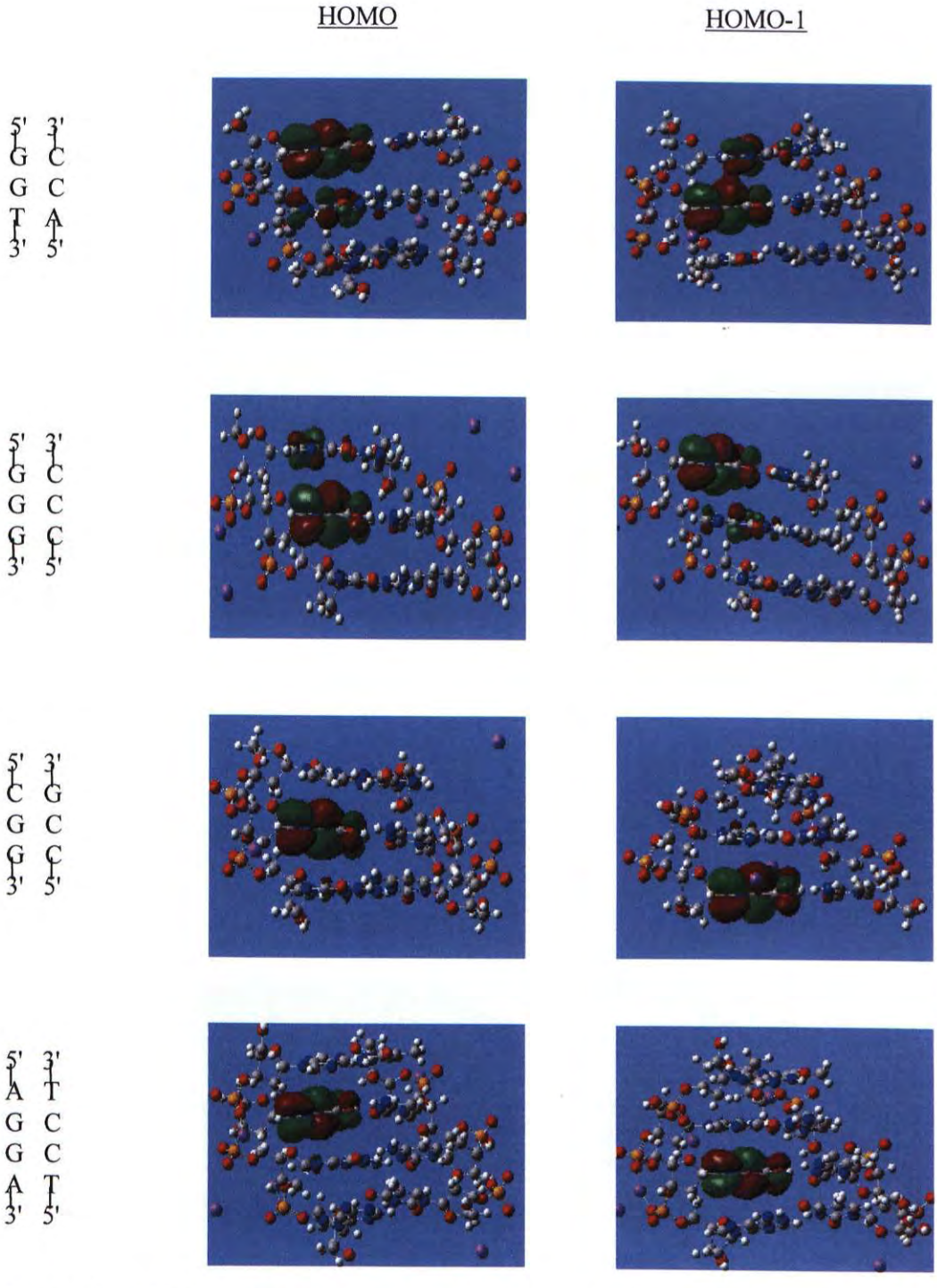
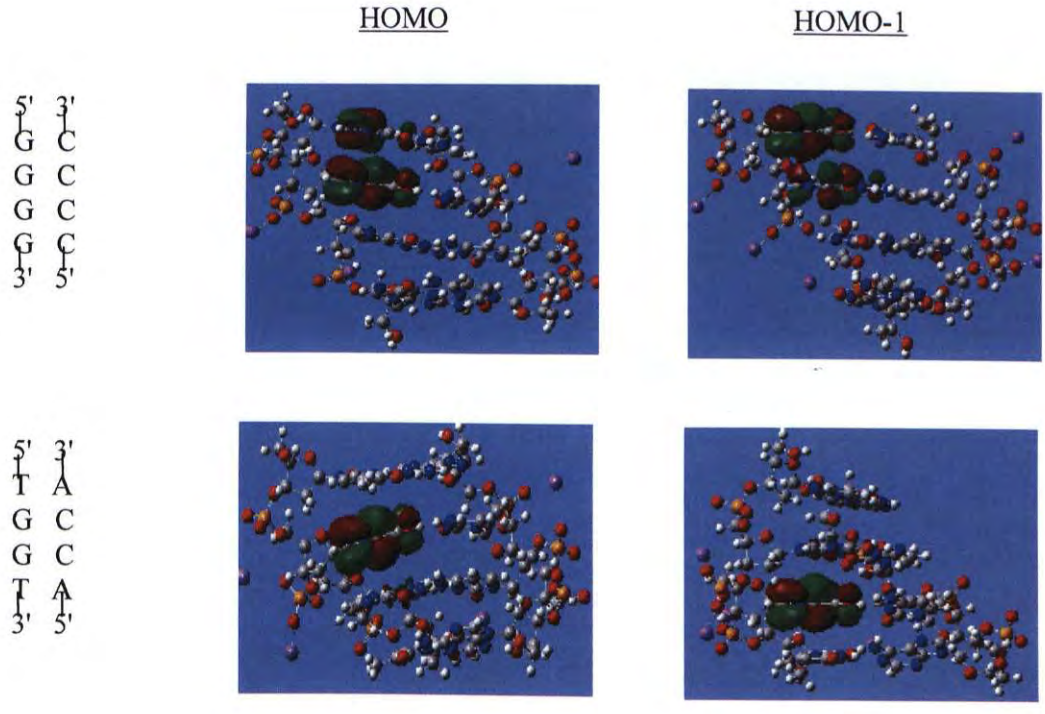


Figure 3.1 The Location of HOMO and HOMO-1 of the Base-pair Models, and the possible Crosslink to Cisplatin

(a) 1,2-d(GpG) Intrastrand Crosslink (cont'd)



(b) 1,2-d(ApG) Intrastrand Crosslink (cont'd)

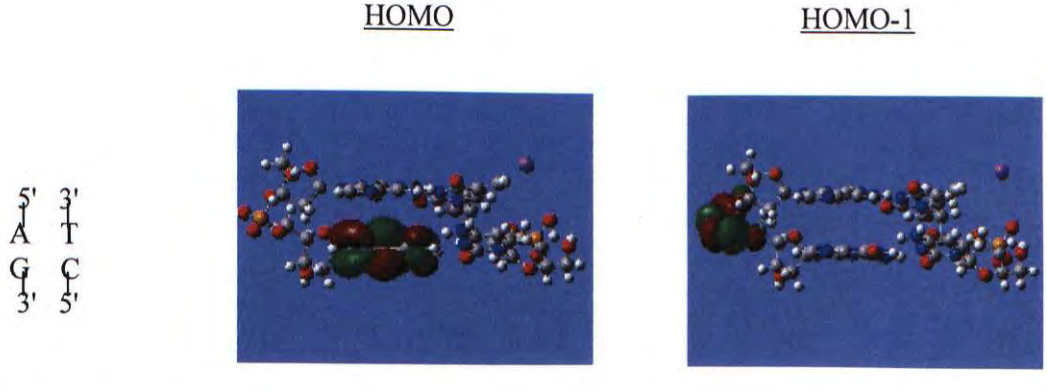
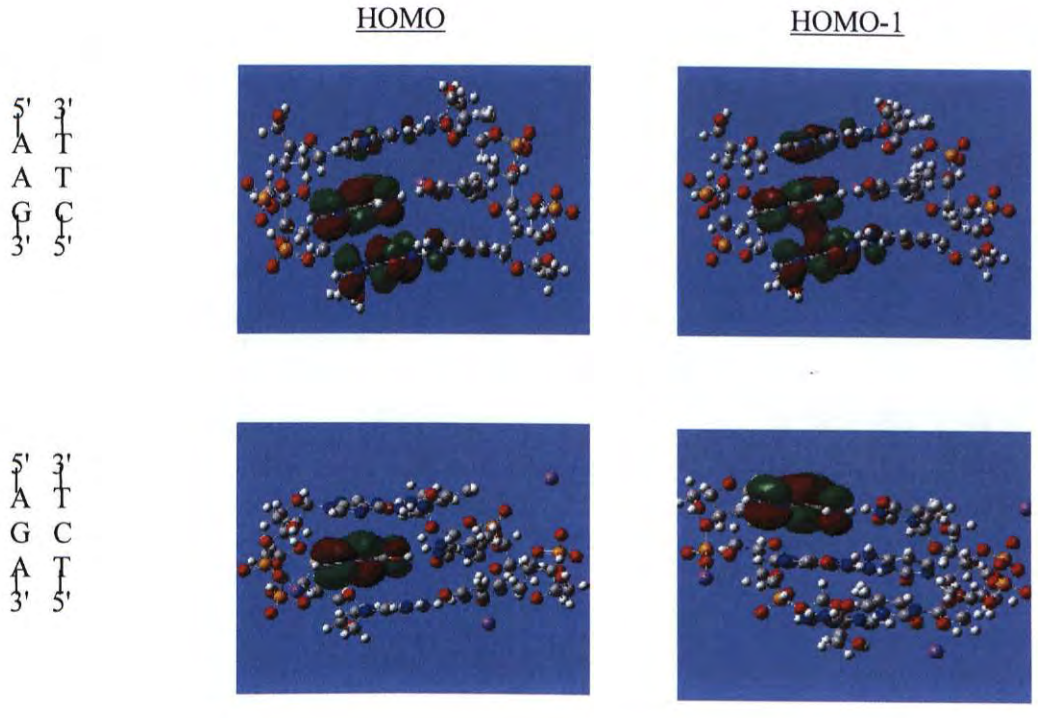


Figure 3.1 The Location of HOMO and HOMO-1 of the Base-pair Models, and the possible Crosslink to Cisplatin

(b) 1,2-d(ApG) Intrastrand Crosslink (cont'd)



(c) 1,3-d(GpXpG) Intrastrand Crosslink (cont'd)

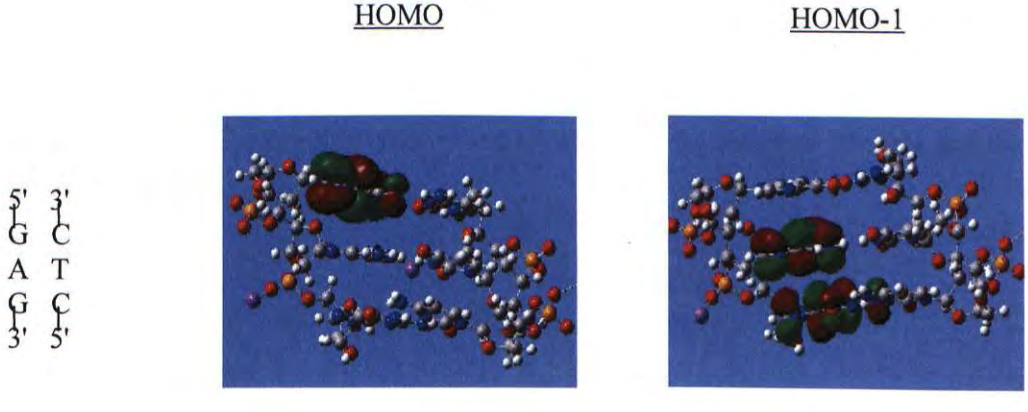


Figure 3.1 The Location of HOMO and HOMO-1 of the Base-pair Models, and the possible Crosslink to Cisplatin

(d) d(GpC)d(GpC) Interstrand Crosslink (cont'd)

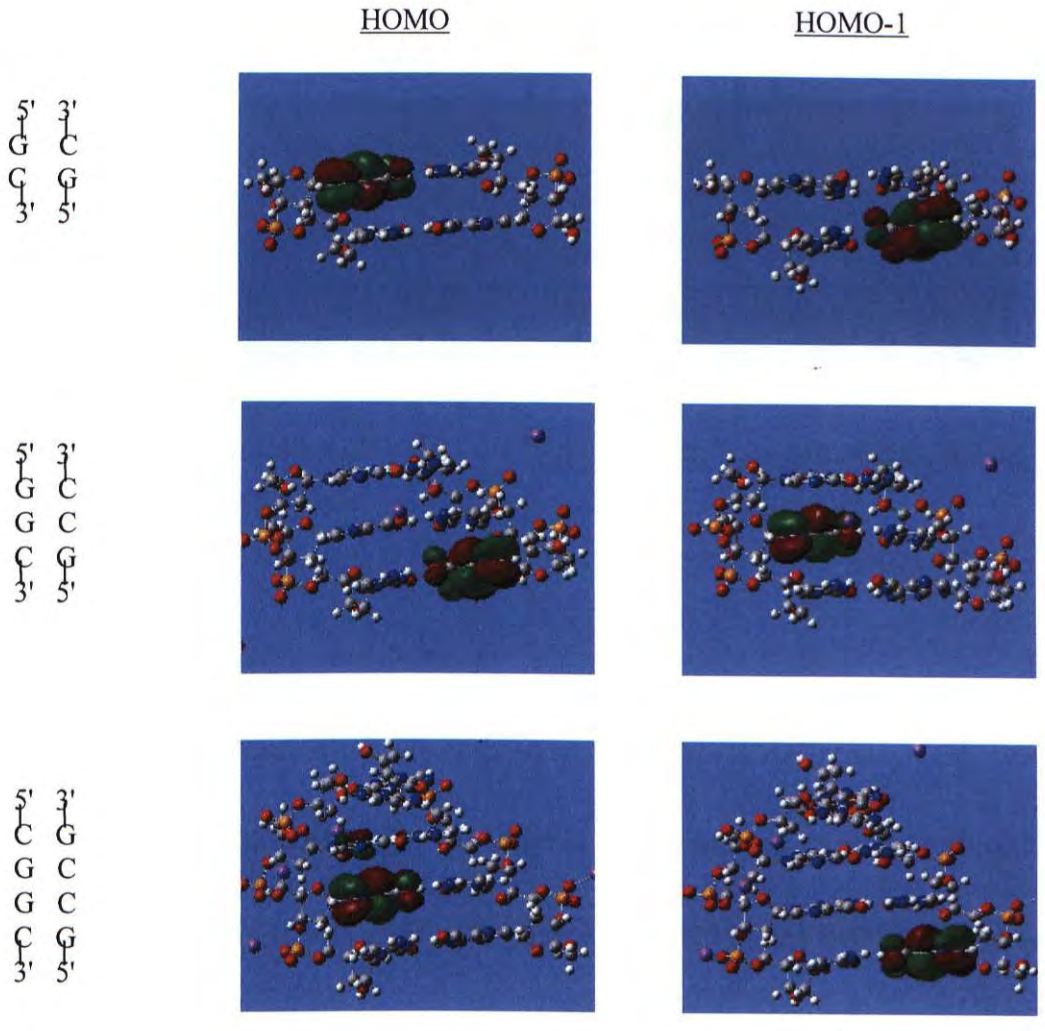


Figure 3.1 The Location of HOMO and HOMO-1 of the Base-pair Models, and the possible Crosslink to Cisplatin

(e) d(GpXpC)d(GpXpC) Interstrand Crosslink (cont'd)

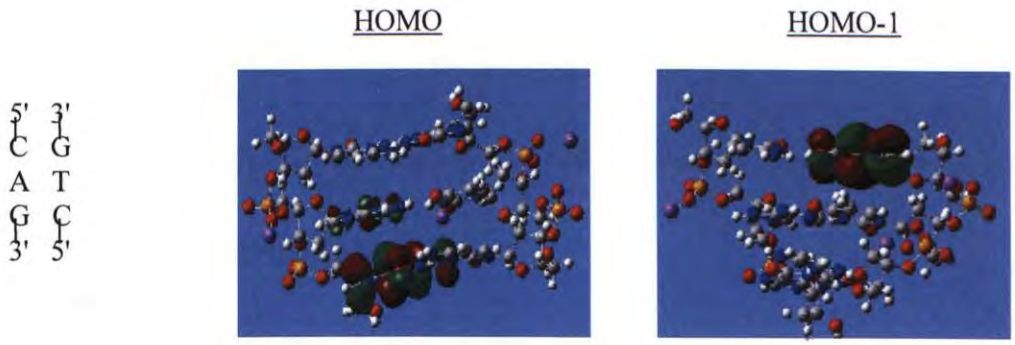


Figure 3.1 The Location of HOMO and HOMO-1 of the Base-pair Models, and the possible Crosslink to Cisplatin

3.3.1 1,2-d(GpG) Intrastrand Crosslink

The HOMOs and HOMO-1s of the models GG, AGG, TGG, GGA, GGT, GGG, CGG, AGGA, GGGG, and TGGT were located on two adjacent Gs (Figure 3.1a). For some models like TGG, GGA, GGT, GGG and GGGG, HOMO and HOMO-1 distributes over two G. However, the ultimate product remains the same.

3.3.2 1,2-d(ApG) Intrastrand Crosslink

The MO calculation results of the DNA models AG, AAG, AGA showed that the potential product with cisplatin was 1,2-d(ApG) intrastrand crosslink (Figure 3.1b). Since there is only one G in the models, the HOMO (or the majority of it) just locates on that G. The HOMO-1s were found on adjacent 5'A of the HOMO-G. This agrees with the

experimental result that 1,2-d(ApG) crosslinks were usual while 1,2-d(GpA) products were hard to find. For the case of AAG, both HOMO and HOMO-1 distribute over two or even three bases, which reveals a fierce competition towards the electrons. However, it can be still concluded that the majority of the products can be determined as 1,2-d(ApG) rather than 1,2-d(ApA) from the figure.

3.3.3 1,3-d(GpXpG) Intrastrand Crosslink

Only the model GAG was determined to be responsible for the 1,3-d(GpXpG) intrastrand crosslink (Figure 3.1c). The HOMO-1 distributes over A and 3'G and this may lead to two products, i.e. 1,2-d(GpA) and 1,3-d(GpXpG). Considering 1,2-d(GpA) is hard to form, 1,3-d(GpXpG) was determined as the ultimate product.

3.3.4 d(GpC)d(GpC) Interstrand Crosslink

The models GC, GGC, CGGC gave the products as the d(GpC)d(GpC) interstrand crosslinks. No obvious electron distribution over bases was observed in these models, which implies high tendency to yield d(GpC)d(GpC) interstrand crosslink.

3.3.5 d(GpXpC)d(GpXpC) Interstrand Crosslink

Only the model CAG gave the product as d(GpXpC)d(GpXpC) interstrand crosslink. Electron distribution profile implies that the tendency to form the unique product is also very high.

3.3.6 Summary

To have a wider and deeper understanding of the MO analysis, the locations of HOMO and HOMO-1 for all re-calculated DNA models are summarized in Figure 3.2.

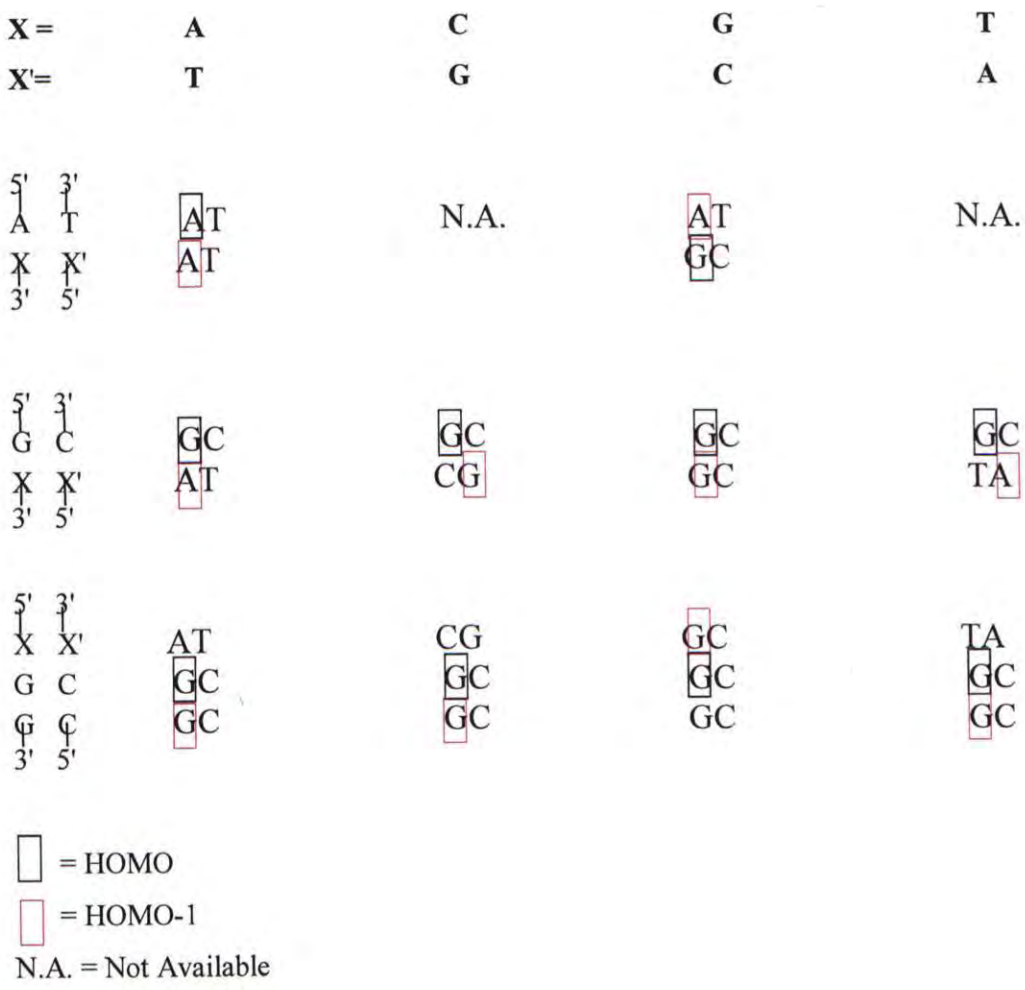


Figure 3.2 Summaries of the Locations of HOMO and HOMO-1 of DNA Models

(cont'd)

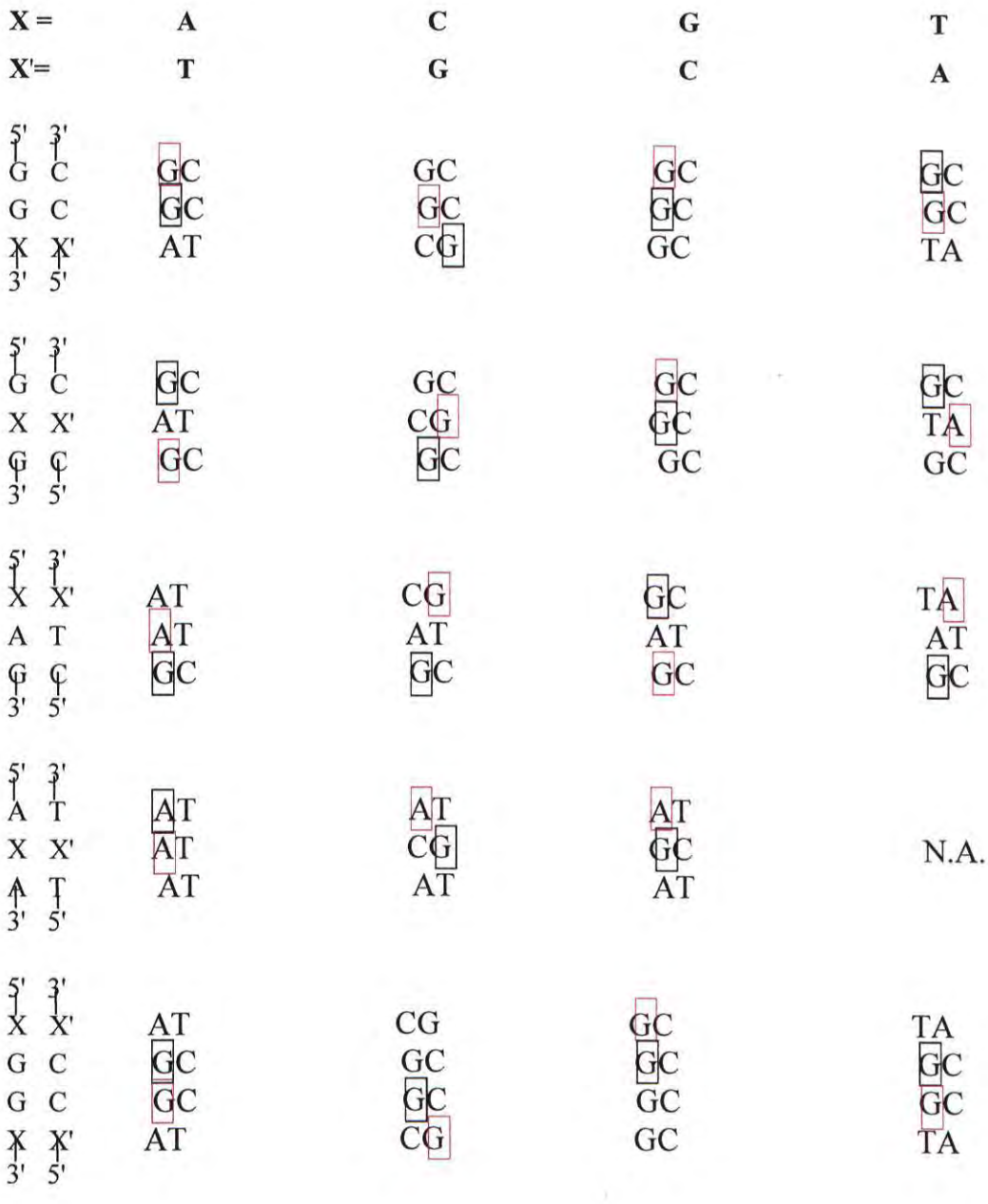


Figure 3.2 Summaries of the Locations of HOMO and HOMO-1 of DNA Models

3.4 Potential Binding Products Analysis

A set of empirical "Selection Rules" was proposed by Fu (18) allowing the HOMO and the nearby active site to be located based on the calculated results. However, as stated in Chapter 2, the methods she adopted suffer from some disadvantages and thus the results obtained were not accurate enough. Here, based on the re-calculated results using the hybrid MD+QM method, a new rule was proposed to predict the potential binding products of DNA with the ligand. The active sites of some experimental Pt bound sequences, which are extracted from Protein Data Bank (PDB), were then predicted using the new rule and a comparison between the two results was carried out.

3.4.1 Site Identification Convention

To make the expression clear, a site identification convention for DNA was given. For a double helix $d(X_2X_1ZY_1Y_2)d(Y_2'Y_1'Z'X_1'X_2')$, on the assumption that strand II is complementary to strand I, the identification detail was shown in Figure 3.3. X_1 and Y_1 are the adjacent neighbours of Z; X_2 and Y_2 are the next adjacent neighbours of Z; X_1' and Y_1' are the nearest neighbours of Z; X_2' and Y_2' are the next nearest neighbours of Z.

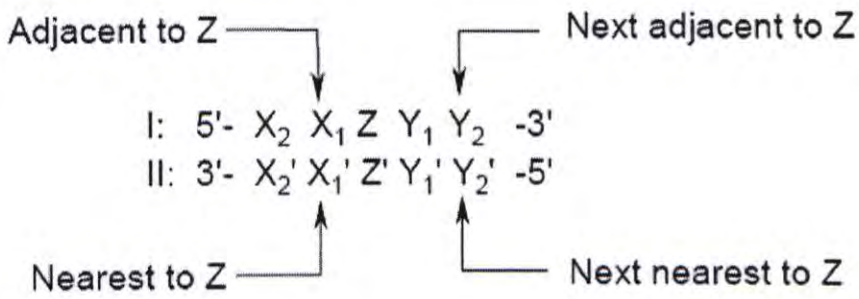


Figure 3.3 Site Identification Conventions

3.4.2 Potential Binding Products Analysis

According to the ligand substitution reaction theory, binding products were resulted from the interaction between the HOMO (and HOMO-1 for bifunctional products) of the ligand and the LUMO of the substrate. The MO analysis above also revealed that the binding of DNA with the ligand is not random but well-regulated. For example, for any sequence containing only one G, the HOMO simply always locates on that G. This is due to the lowest IP value of G among the four DNA bases (Table 1.1). For a sequence containing two or more Gs, the situation becomes complicated. In most cases especially in the case that a sequence containing adjacent purine bases, it just cannot specify the unique location where the HOMO locates due to the fierce competition towards electrons. The ultimate shape of HOMO looks more "diffuse" than that of those sequences containing only one G. But from another point of view, one of the two sites

which compete for the electrons is HOMO, or HOMO-1. So when prediction was to be made, it is more reasonable to use the two active sites together than separately to predict the ultimate products. In other words, the old selection rule which uses the HOMO and HOMO-1 binding site separately to predict the binding products can be replaced by a new advanced rule which uses the two most active binding sites together to directly predict the binding products. The new rule was summarized in Table 3.1 as follows.

Table 3.1 Procedures to Predict the Binding Products of DNA with the Ligand

I. HOMO locates at G for sequences containing only one G. The according potential product was monofunctional, or 1,2-d(ApG).

II. For sequences containing more than one G, HOMO also locates at G. And the potential product can be predicted by the following priority and step outlined.

Step	Details	Potential Binding Products#
1	If there exists such a G with adjacent G or A, or next adjacent G, or nearest G, or next nearest G in the sequences, remove those G(s) without any of them. If n* > 4, also remove terminal G(s).	1,2-intra, 1,3-intra, 1,2-inter, 1,3-inter, 1,2-AG
2	If n > 4, identify G without G or A as adjacent neighbor. If the potential product is inter-strand, identify 5' position G in the other strand.	1,3-intra, 1,2-inter, 1,3-inter

3	If there exists GG in the strand, remove other G(s) only with adjacent A, or next adjacent, nearest, next nearest G or A. If there exist two or more GG, identify that with the most nearest G or A (see GG as a whole). If there are still two or more GG, remove that with the most next adjacent G or A (see GG as two individuals here, and thus there are 4 next adjacent bases in total). If the sequence contains GGC fragment, it's also very possible to form d(GpC)d(GpC) product.	1,2-intra, 1,2-inter
4	1,3-intrastrand	1,3-intra
5	1,2-interstrand	1,2-inter
6	1,3-interstrand	1,3-inter
7	1,2-d(ApG)	1,2-AG

*n represents the number of G in the sequence

#1,2-intra = 1,2-d(GpG); 1,3-intra = 1,3-d(GpXpG); 1,2-inter = d(GpC)d(GpC); 1,3-inter = d(GpXpC)d(GpXpC); 1,2-AG = 1,2-d(ApG)

For sequences containing more than one G, the first thing to do is to distinguish different kinds of G. If a G is totally isolated (i.e. without adjacent G or A, or next adjacent, nearest, next nearest G in the sequence) while other G(s) are not, it's more possible for the ligand to bind with other G(s) to form bifunctional product rather than binding with the isolated G to form monofunctional product. So step 1 is given. Then identify G without G or A as adjacent neighbor. The G(s) identified from this step has no adjacent G or A. This guarantees no fierce competition towards electrons with the G(s) and thus reasonable. The condition $n > 4$ makes sure that the DNA chain is long enough to make sense. Besides, G on the 5' side is believed to be more reactive in general (15, 16). So if the potential product is inter-strand, identify 5' position G in the other strand.

The details in step 3 are a little complicated. This step deals with the situation that there is one or more GG fragment(s) in the sequence. Usually this kind of DNA models can easily form 1,2-d(GpG) crosslink product. So firstly those G(s) only with adjacent A, or next adjacent, nearest, next nearest G or A are removed. If there exist two or more GG, identify that with the most nearest G or A (see GG as a whole). This is because the nearest G or A can help to compete for electrons with its opposite base and thus make the competition between GG and its adjacent base (i.e. the opposite base to the nearest G or A) weaker. If there are still two or more GG left, remove that with the most next adjacent G or A since the next adjacent G or A would compete for electrons with GG fragment and make its IP value become larger and thus electron donation property weaker. In the sequences containing GGC fragment, the statistical result shows that there is also great possibility to form d(GpC)d(GpC) product.

The following procedures are relatively simple. During step 4 1,3-intrastrand crosslink product is formed; then 1,2-interstrand crosslink product formed, step 5; and then 1,3-interstrand crosslink product formed, step 6; and at last 1,2-d(ApG) crosslink product was given. These procedures also come from the statistical results and in most situations it is right. It seems reasonable. One apparent fact is that 1,3-intrastrand crosslink product and 1,2-interstrand crosslink product are easily formed than 1,3-

interstrand crosslink product due to the structural factor. 1,2-d(ApG) are lastly formed due to the thermodynamics factor that the IP value of A (8.24 eV) is higher than that of G (7.75 eV). The difference makes A not that "preferred" by the most active site.

The new rule was then applied to predict the potential binding products of the DNA models GGG, CGG and GGC. They are the very three models the ultimate binding products of which are different from those calculated in gas phase. According to step 3 in Table 3.1, GGG and CGG are determined to give a 1,2-intrastrand product while GGC 1,2-intrastrand or 1,2-interstrand product. This is consistent with the calculated results where GGG and CGG form 1,2-intrastrand products and GGC forms 1,2-interstrand product. In gas phase, however, 1,3-intrastrand product was given by GGG, while 1,2-interstrand product given by CGG and 1,3-interstrand product given by GGC. By comparing the results in water and in gas phase, it can be seen that the results in water are more reasonable. Take GGG as an example, it's more convinced to form a 1,2-intrastrand product than 1,3-intrastrand. Actually for the sequence ATGGGT, the experimental result is indeed the 1,2-intrastrand product, not 1,3-intrastrand product (18).

3.4.3 Applications

To further validate the new selection rule, sequences from the PDB and literature results which had been shown binding with cisplatin were chosen and compared. The results are listed in Table 3.2.

Table 3.2 Comparison of the Predicted versus Experimental Binding Position of GG in the Formation of Cisplatin Binding Adducts

No.	PDB ID	strand	Sequence (5'-xxx–3')		
			Experimental result	Predicted binding site [1]	Predicted binding site [2]
1	1A2E (X-ray)	I II	CCTCGCTCTC GGAGCGAGAG	[1] CCTCGCTCTC GGAGCGAGAG [2]	CCTCGCTCTC GGAGCGAGAG
2	1A84 (NMR)	I II	CCTCTGGTCTCC GGAGACCAGAGG	[1] CCTCTGGTCTCC GGAGACCAGAGG [3] [2]	CCTCTGGTCTCC GGAGACCAGAGG
3	1AIO (X-ray)	I II	CCUCTGGTCTCC GGAGACCAGAGG	[1] CCUCTGGTCTCC GGAGACCAGAGG [3] [2]	CCUCTGGTCTCC GGAGACCAGAGG
4	1AU5 (NMR)	I II	CCTGGTCC GGACCAGG	[1] CCTGGTCC GGACCAGG [3] [2]	CCTGGTCC GGACCAGG
5	1CKT (X-ray)	I II	CCUCTCTGGACCTTCC GGAGAGACCTGGAAGG	[1] CCUCTCTGGACCTTCC GGAGAGACCTGGAAGG [4] [2] [3]	CCUCTCTGGACCTTCC GGAGAGACCTGGAAGG

(cont'd)

	PDB ID	strand	Sequence (5'-xxx–3')		
			Experimental result	Predicted binding site [1]	Predicted binding site [2]
6	1DDP (NMR)	I II	CATAGCTATG GTATCGATAC	CATAGCTATG GTATCGATAC	CATAGCTATG GTATCGATAC
7	1I1P (X-ray)	I II	CCTCGCTCTC GGAGCGAGAG	[1] CCTCGCTCTC GGAGCGAGAG [2]	CCTCGCTCTC GGAGCGAGAG
8	1KSB (NMR)	I II	CTCCGGCCT GAGGCCGGA	[1] CTCCGGCCT GAGGCCGGA [2] [3]	CTCCGGCCT GAGGCCGGA
9	2NPW (NMR)	I II	CCTCAGGCCTCC GGAGGCCTGAGG	[1] CCTCAGGCCTCC GGAGGCCTGAGG [3] [4] [2]	CCTCAGGCCTCC GGAGGCCTGAGG
10	3CO3 (X-ray)	I II	CCTCTCGTCTCC GGAGACGAGAGG	CCTCTCGTCTCC GGAGACGAGAGG	CCTCTCGTCTCC GGAGACGAGAGG
11	&	I II	TATGCATA ATACGTAT	TATGCATA ATACGTAT	TATGCATA ATACGTAT
12	&	I II	TATGGCAT ATACCGTA	[2] TATGGCAT ATACCGTA [1]	TATGGCAT ATACCGTA

(cont'd)

	PDB ID	strand	Sequence (5'-xxx–3')		
			Experimental result	Predicted binding site [1]	Predicted binding site [2]
13	&	I II	TATGGTAT ATACCATA	TATGGTAT ATACCATA	TATGGTAT ATACCATA
14	&	I II	...CTGTGCA... ...GACACGT...	[1] ...CTGTGCA... ...GACACGT... [2]	...CTGTGCA... ...GACACGT...
15	&	I II	CTCTCGGTCTC GAGAGCCAGAG	CTCTCGGTCTC GAGAGCCAGAG	CTCTCGGTCTC GAGAGCCAGAG
16	&	I II	ATACATGGTACATA TATGTACCATGTAT	ATACATGGTACATA TATGTACCATGTAT	ATACATGGTACATA TATGTACCATGTAT
17	&	I II	ATGGGT TACCCA	ATGGGT TACCCA	ATGGGT TACCCA

Number with square bracket [ ] indicates the priority; [1] indicates the results obtained in gas phase; [2] indicates the results obtained in water; & indicates no PDB ID number but literature results (18).

It's known from the table that the predicted binding sites using the new selection rule match the experimental results perfectly. As comparison, there is too much uncertainty for the predicted binding sites using the "old" selection rule (18). It indicates the new rule was more practicable and applicable for the realistic situation. The only

model which detains some uncertainty under the new rule is sequence No. 17. There is a GGG fragment in the sequence. Generally speaking, the new rule can't deal with three or more consecutive Gs situation. It's necessary to have an extra calculation towards pure G models and the results can be used specially to predict that situation. The binding results for 5-8 consecutive Gs are: GGG, GGGG, GGGGG, GGGGGG, GGGGGGG, GGGGGGGG. That three Gs were underlined at the same time in some models indicates the electrons of HOMO and HOMO-1 diffuse over those three bases and hard to distinguish. Two different kinds of 1,2-intrastrand crosslink products may thus be formed.

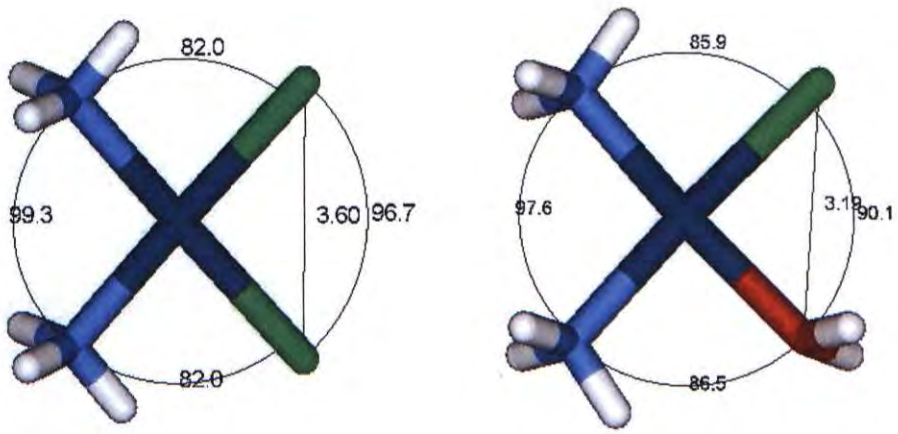
3.5 Cisplatin-DNA Crosslink Products Analysis

As an application of the calculated results, cisplatin as an important ligand binding with DNA was carefully examined structurally.

The chemical structural formula of cisplatin (CDDP) was listed in Figure 1.10a. Cisplatin and cisplatin-H_2O (CDDP-H_2O) was ever optimized at the RHF/STO-3G and RB3LYP/LANL2DZ level of theory (18). Figure 3.4 illustrates the optimized structures of cisplatin (a) and cisplatin- H_2O (b) at RB3LYP/LANL2DZ level of theory.

In theory five crosslink products can be formed (chapter 3.3.1-3.3.5). In experiments however, only four products had been identified. They are 1,2-d(GpG), 1,2-

d(ApG), 1,3-d(GpXpG) intrastrand and d(GpC)d(GpC) interstrand crosslink adducts (62-64). In order to further understand the interaction between cisplatin and DNA, the helical parameters of the optimized geometries and experimental sequences were analyzed with CURVES 4.1 (65-67). Analysis will be firstly focused on the inter base-pair parameters, namely shift (Dx), slide (Dy), rise (Dz), tilt (τ), roll (ρ) and twist (Ω) (Figure 1.8). Then the four experimental crosslink products will be discussed in details. The torsion angles of the sugar-phosphate backbone (α, β, γ, δ, ε and ξ) and the orientation of the base (χ) (Figure 1.6), together with the sugar ring conformation of each structure were also adopted to facilitate the analysis (Appendix I).



(a) Cisplatin (b) Cisplatin-H_2O

Figure 3.4 Diagrams of Optimized (a) Cisplatin and (b) Cisplatin-H_2O at RB3LYP/LANL2DZ Level of Theory

Detailed results of the inter base-pair parameters for all optimized geometries are summarized in Table 3.3. Table 3.4 lists the average value from the computed base-pair structures for each base-pair step.

Table 3.3 Inter Base Pair Parameters of the Optimized Geometries [a]

Base Pair	Step	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (τ)	Roll (ρ)	Twist (Ω)
(AX, TX')							
(AA,TT)	A1/A2	0.02	0.00	3.22	0.13	0.28	33.17
(AG,TC)	A1/G2	0.06	-0.05	3.19	-0.86	1.24	32.67
(GX, CX')							
(GA,CT)	G1/A2	0.02	-0.08	3.40	-0.75	1.53	37.20
(GC,CG)	G1/C2	0.03	-0.02	3.77	-1.43	1.06	34.75
(GG,CC)	G1/G2	0.00	-0.02	3.75	0.32	-0.02	32.08
(GT,CA)	G1/T2	-0.02	-0.01	3.11	1.12	0.25	31.86
(XGG, X'CC)							
(AGG,TCC)	A1/G2	-0.38	0.24	2.64	-0.44	4.09	30.34
	G2/G3	0.18	-0.34	3.91	-1.57	-2.08	32.99
(CGG,GCC)	C1/G2	-0.64	-0.03	2.98	-4.07	2.44	37.29
	G2/G3	0.68	-0.23	3.59	-2.32	-4.77	35.26
(GGG,CCC)	G1/G2	0.15	0.04	3.77	0.98	-0.63	30.98
	G2/G3	-0.10	-0.04	3.64	-1.78	1.54	30.47
(TGG,ACC)	T1/G2	0.55	-0.28	2.71	0.55	16.10	25.95
	G2/G3	-0.41	0.17	3.66	-1.12	-2.83	38.95

(cont'd)

Base Pair	Step	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (τ)	Roll (ρ)	Twist (Ω)
(GGX, CCX')							
(GGA,CCT)	G1/G2	0.16	-0.47	3.74	3.85	1.45	34.44
	G2/A3	0.00	0.47	3.07	1.19	3.49	38.00
(GGC,CCG)	G1/G2	-0.26	0.01	3.33	2.41	1.09	34.29
	G2/C3	0.24	-0.15	3.33	-2.42	-0.17	33.59
(GGT,CCA)	G1/G2	-0.51	0.36	3.34	-4.81	1.98	37.48
	G2/T3	0.43	-0.69	3.33	-2.13	-1.93	31.54
(GXG, CX'C)							
(GCG,CGC)	G1/C2	0.11	-0.33	3.42	2.12	-5.38	33.08
	C2/G3	0.01	0.27	2.68	2.57	10.21	30.58
(GTG,CAC)	G1/T2	-0.43	-0.37	3.00	2.13	-4.36	33.06
	T2/G3	0.48	0.17	3.02	1.51	2.81	38.15
(XAG, X'TC)							
(AAG,TTC)	A1/A2	-0.71	0.38	3.17	-7.77	-2.47	35.88
	A2/G3	0.72	-0.91	3.33	-3.49	5.31	28.49
(CAG,GTC)	C1/A2	-0.42	0.28	3.04	0.18	3.15	36.29
	A2/G3	0.38	-0.52	2.92	-4.69	-3.04	30.30
(GAG,CTC)	G1/A2	-0.32	-0.05	3.11	0.11	0.02	37.09
	A2/G3	0.33	-0.25	3.17	-0.76	4.32	30.13
(TAG,ATC)	T1/A2	-0.36	0.21	2.90	0.12	2.81	35.42
	A2/G3	0.24	-0.51	2.67	-1.48	-0.29	27.95
(AXA, TX'T)							
(AAA,TTT)	A1/A2	0.02	-0.21	3.28	1.65	1.55	33.96
	A2/A3	-0.07	0.04	2.86	2.69	3.18	33.36

(cont'd)

Base Pair	Step	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (τ)	Roll (ρ)	Twist (Ω)
(AGA,TCT)							
	A1/G2	-0.06	-0.05	3.35	-1.32	7.13	35.93
	G2/A3	0.13	-0.11	3.60	-4.97	-3.60	35.80
(XGGX, X'CCX')							
(AGGA,TCCT)							
	A1/G2	-0.35	0.14	3.15	-1.61	-0.08	35.35
	G2/G3	0.05	-0.96	3.43	3.07	0.73	29.15
	G3/A4	0.40	0.43	3.05	0.67	2.14	37.08
(CGGC,GCCG)							
	C1/G2	-0.49	0.36	2.63	-3.48	4.96	33.93
	G2/G3	0.52	-0.39	3.18	-4.34	-0.08	28.32
	G3/C4	-0.22	-0.26	3.13	-1.74	0.39	29.31
(GGGG,CCCC)							
	G1/G2	0.16	-0.02	3.60	3.16	-1.78	32.46
	G2/G3	-0.40	-0.30	3.29	1.73	2.60	30.05
	G3/G4	0.29	0.21	2.69	1.10	3.65	29.83
(TGGT,ACCA)							
	T1/G2	0.26	-0.13	2.53	-2.10	17.91	25.32
	G2/G3	-0.53	0.27	3.93	-4.49	-7.88	39.96
	G3/T4	0.23	-0.51	3.41	0.29	1.28	31.31

[a] Distances (Dx, Dy, Dz) are given in angstroms, and angles (τ,ρ,Ω) are in degrees.

Table 3.4 Summary of the Inter Base-Pair Parameters of Computed Base-Pair Step[a]

Base-Pair Step	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (τ)	Roll (ρ)	Twist (Ω)
A/A	-0.18	0.05	3.13	-0.82	0.64	34.09
A/G	0.12	-0.24	3.05	-1.84	2.34	31.40
G/A	0.05	0.13	3.25	-0.75	0.72	37.03
G/C	0.04	-0.19	3.41	0.87	-1.02	32.68

(cont'd)

Base-Pair Step	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (τ)	Roll (ρ)	Twist (Ω)
G/T	0.05	0.40	3.21	0.35	-1.19	31.94
G/G	0.00	-0.11	3.52	-0.25	-0.47	33.11
C/G	-0.37	0.20	2.76	1.66	5.87	33.93
T/A	-0.36	0.21	2.90	0.12	2.81	35.42
T/G	0.43	-0.08	2.75	-0.01	12.27	29.81

[a] Distances (Dx, Dy, Dz) are given in angstroms, and angles (τ,ρ,Ω) are in degrees.

3.5.1 1,2-d(GpG) and 1,2-d(ApG) Intrastrand Crosslinks

For these two intrastrand products, the HOMO and HOMO-1 locates on the adjacent purines. From Table 3.4 it's known that the average rises for step AG and GG are 3.05, 3.52, respectively. It's reported by Astbury (3) and Wilkins (68) that the stacked base-pair distance from the X-Ray data of DNA is 3.4 Å. All the three data closely approaches the chloride to chloride distance, r(Cl-OH_2), of 3.19 Å of the optimized cisplatin-H_2O (Figure 3.4). Thus the binding to cisplatin is feasible.

For 1,2-d(ApG) and 1,2-d(GpA) intrastrand crosslinks, both the MO results (Figure 3.2) and experiments (28) show that binding of cisplatin takes place on the A of the 5' side of d(ApGpA). The reason remains unclear (69). Considering that a larger rise and a larger twist angle for GA step (3.25, 37.03, respectively) than the value of AG step

(3.05, 31.40, respectively) (Table 3.4), it would be more difficult for ligation of GA to cisplatin.

3.5.2 1,3-d(GpXpG) Intrastrand and d(GpXpC)d(GpXpC) Interstrand Crosslinks

The MO results show that only GAG model forms 1,3-d(GpXpG) intrastrand crosslink product and CAG d(GpXpC)d(GpXpC) interstrand crosslink product. The N7-N7 distances of these two models were measured and summarized (Table 3.5). Both the optimized and system-built distances are much longer than r(Cl-OH_2) of 3.19 Å. Even though a distance of 6.8 Å [2 x 3.4Å based on Astbury's (3) and Wilkins' (68)] was predicted for double stacked base-pairs, it remains much longer. Thus large structural distortion on DNA would happen if binding with cisplatin occurs (15, 16), which renders these crosslinks structurally less favorable.

Table 3.5 N7-N7 Distance of G-G of the Optimized & System-Built GAG and CAG

Model		GAG	CAG
N7-N7 Distance /Å	Optimized	7.96	10.31
	AMBER	8.18	9.44

3.5.3 d(GpC)d(GpC) Interstrand Crosslinks

GC, GGC, CGGC models are showed to give the d(GpC)d(GpC) interstrand crosslinks according to the MO results. The N7-N7 distances of these three models were measured and summarized (Table 3.6). It's known from the table that the optimized N7-N7 distance is 8.69 Å around. This value is larger than that of 1,3-d(GpXpG) intrastrand crosslinks (7.96 Å, Table 3.5) but smaller than that of d(GpXpC)d(GpXpC) interstrand crosslinks (10.31 Å, Table 3.5). It indicates that it's relatively hard to form d(GpC)d(GpC) products than 1,3-d(GpXpG) intrastrand crosslinks, but easier than d(GpXpC)d(GpXpC) interstrand crosslinks. The conclusion therefore validates parts of the new rule that 1,3-intrastrand products firstly formed, then 1,2-interstrand and then 1,3-interstrand (Table 3.1).

It's also interesting to note that the three N7-N7 values of system-built GC, GGC and CGGC models are all equal to 7.77 Å. Since CGGC has one more 5'-C than GGC and GGC has one more 5'-G than GC, the observation may imply that when AMBER software builds models, it just "adds" the base one by one without changing the conformation of the former structure. This hypothesis was proved by comparing other inter base-pair parameters of these three system-built models (Table 3.7). It was found that each parameter of the three models is almost equal from one another.

Table 3.6 N7-N7 Distance of the Optimized & System-Built GC, GGC and CGGC

Model		GC	GGC	CGGC
N7-N7 Distance /Å	Optimized	8.76	8.67	8.63
	AMBER	7.77	7.77	7.77

Table 3.7 Inter Base Pair Parameters of the System-Built GC, GGC and GGGC [a]

Base Pair	Step	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (τ)	Roll (ρ)	Twist (Ω)
GC	G1/C2	0.00	0.00	3.39	0.00	0.00	36.00
GGC	G1/G2	0.00	0.00	3.38	0.02	-0.04	36.00
	G2/C3	0.00	0.00	3.39	-0.01	0.05	35.98
CGGC	C1/G2	0.00	0.00	3.37	0.01	-0.09	36.01
	G2/G3	0.00	0.00	3.38	0.03	0.00	35.98
	G3/C4	0.00	0.00	3.39	0.01	0.09	36.03

[a] Distances (Dx, Dy, Dz) are given in angstroms, and angles (τ,ρ,Ω) are in degrees.

However, even the smallest value of N7-N7 distance in Table 3.6 is much larger than r(Cl-OH_2) of 3.19 Å. So like the situation of d(GpXpG) intrastrand crosslink, large structural distortion on DNA would be induced if binding with cisplatin occurs, which also renders these crosslinks structurally less favorable.

3.5.4 Platination at Terminal Positions

Generally cisplatin would not bind with terminal bases to form cisplatin-DNA adduct (Table 3.2) even if GG or AG fragment locates at terminal position. We believe this is attributed to the terminal effect. The binding of cisplatin would become increasingly difficult due to the increase latitude of the "moving" nucleotides in the terminal position. This conclusion can be rationalized by comparing the backbone torsion angles among the results obtained from the calculation, NMR and X-ray trials. The experimental sequences were obtained from the Nucleic Acid Database (NDB) (49) where all sequences terminated with GGG were collected (Appendices I and II) (70). It's known that the angles of the terminal base deviate more or less from the experimental results, even NMR and X-ray data deviate from each other (Figure 2.2-2.5). However for the intermediary base this circumstance does not exist, confirming that the terminal base is dynamically unstable.

3.6 Summary

In this chapter, both the orbital and structures of the DNA models were re-calculted using the hybrid MD+QM method. According to the MO results, a new selection rule was proposed. Different from the "old" selection rule which locates the

HOMO and HOMO-1 in DNA separately, the new rule locates the two most active sites together to predict the potential binding products. The application of the new rule to the experimental sequences was extremely successful. Then cisplatin-DNA crosslink products were carefully examined from the structural point of view. The analysis not only further proved the correctness of the new rule, but also made the understanding towards the binding of cisplatin and DNA deeper.

As a conclusion, the preference of binding is found to be sequence dependent, and both the electronic and structural factors heavily influence the ligation of DNA to cisplatin.

CHAPTER FOUR

CONCLUDING REMARKS

The traditional hybrid Molecular Dynamics + Quantum Mechanics (MD+QM) method was utilized to theoretically investigate the potential crosslink products of DNA with cisplatin. A series of short segment DNA models was selected. They are 5'-GX-3', 5'-AX-3', 5'-XGG-3', 5'-GGX-3', 5'-GXG-3', 5'-XAG-3', 5'-AXA-3'and 5'-XGGX-3' (where X = A, T, C and G). These models were firstly involved in a 120 ps MD simulation process, and then their geometries were averaged over the simulation time during which the Rmsd of the models has no wild fluctuation. Lastly MO analysis of their structures was undertaken at the RHF/GEN (6-31G* for P atom, 3-21G for other atoms) level of theory.

Three parameters, namely backbone torsion angles, the reference distance as well as the HOMO location were compared between the data obtained from the hybrid method and the experimental data, or the results obtained from the reliable computational method. All of the results are highly compatible, and much better than the results obtained from the pure quantum mechanics method RHF/STO-3G which has been proved to be a

qualitative method for the short segment DNA models, indicating that the utilized hybrid method was highly appropriate for the demonstration of the investigation.

MO analysis was then carried out on all the DNA models, based on which a new selection rule was proposed. Different from the "old" selection rule which locates the HOMO and HOMO-1 in DNA separately, the new rule locates the two most active sites together to predict the potential binding products. The application of the new rule to the experimental sequences was extremely successful.

Then the four experimental crosslink products, namely 1,2-d(ApG), 1,2-d(GpG) and 1,3-d(GpXpG) intrastrand and d(GpC)d(GpC) interstrand crosslinks were carefully examined structurally. The results are in general agreement with experimental ones, although the different binding status, namely one bound, one not.

The results also have some application on other anti-cancer drugs like intercalating-type drugs. These successful applications to the practical issues as well as the low computational cost consolidate the method utilized in this thesis to be a good choice when dealing with DNA and other biomacromolecule.

APPENDIX I

BACKBONE TORSION ANGLES AND SUGAR RING CONFORMATIONS OF THE OPTIMIZED GEOMETRIES

	α P-O5'	β O5'-C5'	γ C5'-C4'	δ C4'-C3'	ε C3'-O3'	ξ O3'-P	χ C1'-N	Conformation
(AX,TX')								
(AA,TT)								
5'AA								
A1	-78.73	163.76	58.52	131.61	-143.94	-80.66	-138.36	C2'-endo
A2	/	/	49.04	122.69	/	/	-97.39	C1'-exo
5'TT								
T1	-75.10	173.65	50.13	129.57	-162.65	-83.89	-133.01	C1'-exo
T2	/	/	46.88	119.32	/	/	-111.52	C1'-exo
(AG,TC)								
5'AG								
A1	-73.62	171.28	58.47	132.23	-171.38	-88.65	-115.86	C2'-endo
G2	/	/	52.07	121.29	/	/	-118.45	C1'-exo
5'CT								
C1	-72.37	176.55	56.29	124.91	-167.11	-84.51	-136.01	C1'-exo
T2	/	/	51.50	124.91	/	/	-115.88	C1'-exo
(GX,CX')								
(GA,CT)								
5'GA								
G1	-82.19	149.43	/	137.99	-78.85	137.93	-72.00	C2'-endo
A2	/	/	39.94	139.62	/	/	-100.01	C2'-endo
5'TC								
T1	-69.41	170.02	58.00	117.15	-166.53	-86.52	-125.30	C1'-exo
C2	/	/	52.80	121.27	/	/	-106.89	C1'-exo
(GC,GC)								
5'GC								
G1	-69.82	176.85	57.06	119.31	-172.95	-87.16	-137.01	C1'-exo

(cont'd)

	α P-O5'	β O5'-C5'	γ C5'-C4'	δ C4'-C3'	ε C3'-O3'	ξ O3'-P	χ C1'-N	Conformation
C2	/	/	54.72	121.45	/	/	-128.25	C1'-exo
5'CG								
C1	/	/	51.98	126.81	/	/	-125.94	C1'-exo
G2	-67.00	-179.39	/	136.53	-172.96	-89.59	-133.86	C2'-endo
(GG,CC)								
5'GG								
G1	-69.23	179.74	57.32	123.62	-170.72	-88.79	-137.49	C1'-exo
G2	/	/	48.44	120.96	/	/	-129.26	C1'-exo
5'CC								
C1	-71.08	174.82	58.18	124.70	-167.18	-83.51	-133.01	C1'-exo
C2	/	/	49.53	120.24	/	/	-130.19	C1'-exo
(GT,CA)								
5'GT								
G1	-67.11	176.60	58.15	95.50	-162.13	-76.06	-159.82	C4'-exo
T2	/	/	60.03	127.50	/	/	-127.71	C1'-exo
5'AC								
A1	-70.75	173.71	57.55	129.16	-169.30	-84.72	-126.21	C2'-endo
C2	/	/	49.07	117.70	/	/	-122.90	C1'-exo
(XGG,X'CC)								
(AGG,TCC)								
5'AGG								
A1	-64.31	174.60	59.06	84.93	-166.62	-75.08	-155.58	C3'-endo
G2	-68.44	174.12	66.25	141.60	-173.59	-88.29	-120.23	C2'-endo
G3	/	/	50.74	114.68	/	/	-132.99	C1'-exo
5'CCT								
C1	-73.64	-178.90	55.62	118.95	-164.33	-85.49	-143.98	C1'-exo
C2	-70.17	170.29	49.70	118.14	-168.51	-86.48	-129.64	C1'-exo
T3	/	/	51.09	113.15	/	/	-121.10	C1'-exo

(cont'd)

	α P-O5'	β O5'-C5'	γ C5'-C4'	δ C4'-C3'	ε C3'-O3'	ξ O3'-P	χ C1'-N	Conformation
(CGG,GCC)								
5'CGG								
C1	-73.98	-178.87	56.10	112.98	-165.65	-98.75	-131.17	C1'-exo
G2	-66.58	170.77	51.02	138.67	-175.96	-91.00	-106.99	C2'-endo
G3	/	/	53.44	122.56	/	/	-117.39	C1'-exo
5'CCG								
C1	-73.79	171.34	62.58	109.22	-159.71	-121.96	-116.85	C1'-exo
C2	-73.48	171.96	52.69	135.06	-169.24	-88.02	-115.82	C2'-endo
G3	/	/	47.26	118.17	/	/	-113.88	C1'-exo
(GGG,CCC)								
5'GGG								
G1	-64.93	171.32	57.76	121.80	-175.69	-91.37	-127.34	C1'-exo
G2	-70.34	179.17	58.15	116.27	-173.33	-88.77	-136.21	C1'-exo
G3	/	/	49.80	118.99	/	/	-133.40	C1'-exo
5'CCC								
C1	-67.82	174.69	57.42	114.50	-170.84	-86.94	-135.42	C1'-exo
C2	-68.22	-178.41	56.13	118.39	-173.11	-91.11	-134.60	C1'-exo
C3	/	/	49.39	122.95	/	/	-133.28	C1'-exo
(TGG,ACC)								
5'TGG								
T1	-75.87	170.01	59.51	123.86	-161.95	-80.60	-123.07	C1'-exo
G2	-82.74	157.05	44.27	137.33	-82.46	152.71	-92.46	C1'-exo
G3	/	/	34.83	144.29	/	/	-109.44	C2'-endo
5'CCA								
C1	-65.54	171.90	56.13	123.86	-174.52	-89.66	-120.00	C1'-exo
C2	-71.34	170.78	56.42	117.22	-169.15	-88.76	-118.10	C1'-exo
A3	/	/	50.05	125.77	/	/	-105.03	C1'-exo
(GGX,CCX')								
(GGA,CCT)								
5'GGA								
G1	-65.91	173.64	56.43	119.78	-173.93	-91.49	-130.16	C1'-exo

(cont'd)

	α P-O5'	β O5'-C5'	γ C5'-C4'	δ C4'-C3'	ε C3'-O3'	ξ O3'-P	χ C1'-N	Conformation
G2	-68.33	-178.84	53.93	107.36	-171.67	-88.47	-129.94	O1'-endo
A3	/	/	50.96	120.33	/	/	-108.83	C1'-exo
5'TCC								
T1	-66.16	169.90	59.90	115.13	-174.62	-91.07	-125.35	C1'-exo
C2	-70.18	174.06	61.79	128.13	-175.86	-94.02	-111.77	C2'-endo
C3	/	/	51.57	109.15	/	/	-131.97	O1'-endo
(GGC,CCG)								
5'GGC								
G1	-64.90	-177.78	57.11	108.75	-171.01	-89.18	-141.50	O1'-endo
G2	-68.15	173.19	53.48	135.95	-173.13	-90.65	-118.81	C2'-endo
C3	/	/	52.55	115.56	/	/	-127.45	C1'-exo
5'GCC								
G1	-72.06	175.19	58.64	112.86	-161.98	-77.50	-153.00	C1'-exo
C2	-66.65	174.74	56.25	124.31	-174.27	-89.20	-126.40	C1'-exo
C3	/	/	51.89	118.09	/	/	-128.72	C1'-exo
(GGT,CCA)								
5'GGT								
G1	-79.08	165.49	58.22	129.28	-100.50	164.19	-93.07	C1'-exo
G2	-67.51	171.92	41.68	142.11	-179.11	-94.52	-105.66	C2'-endo
T3	/	/	55.04	113.56	/	/	-122.88	C1'-exo
5'ACC								
A1	-62.84	172.96	58.01	117.14	-174.70	-89.78	-125.48	C1'-exo
C2	-72.23	170.86	58.00	123.97	-164.41	-86.73	-119.62	C1'-exo
C3	/	/	52.65	128.14	/	/	-99.54	C1'-exo
(GXG,CX'C)								
(GCG,CGC)								
5'GCG								
G1	-69.21	172.21	58.46	126.92	-173.63	-89.90	-119.87	C2'-endo
C2	-65.63	171.15	49.50	93.68	-173.27	-83.55	-129.94	O1'-endo
G3	/	/	54.63	117.10	/	/	-113.12	C1'-exo

(cont'd)

	α P-O5'	β O5'-C5'	γ C5'-C4'	δ C4'-C3'	ε C3'-O3'	ξ O3'-P	χ C1'-N	Conformation
5'CGC								
C1	-75.62	173.28	59.88	101.93	-162.87	-81.71	-135.97	O1'-endo
G2	-67.79	162.65	56.44	135.83	-152.46	-131.67	-99.92	C2'-endo
C3	/	/	52.73	126.84	/	/	-118.93	C1'-exo
(GTG,CAC)								
5'GTG								
G1	-68.77	178.16	56.29	111.62	-165.68	-85.16	-146.06	C1'-exo
T2	-68.48	172.48	56.35	127.87	-173.46	-91.60	-116.78	C2'-endo
G3	/	/	50.75	132.82	/	/	-110.20	C1'-exo
5'CAC								
C1	-69.02	-179.65	58.23	102.11	-164.63	-86.24	-136.49	O1'-endo
A2	-66.43	169.05	53.00	137.44	-175.36	-91.38	-105.45	C2'-endo
C3	/	/	54.64	125.29	/	/	-109.86	C1'-exo
(XAG,X'TC)								
(AAG,TTC)								
5'AAG								
A1	-93.41	158.07	59.65	127.41	-86.07	121.91	-68.10	C1'-exo
A2	-71.89	171.46	57.55	140.55	-172.22	-88.67	-107.27	C2'-endo
G3	/	/	49.31	123.19	/	/	-110.58	C1'-exo
5'CTT								
C1	-65.94	174.99	56.18	106.58	-175.38	-90.44	-133.85	C1'-exo
T2	-69.25	170.51	54.51	108.41	-170.06	-84.66	-127.35	C1'-exo
T3	/	/	55.56	117.50	/	/	-111.35	C1'-exo
(CAG,GTC)								
5'CAG								
C1	-68.18	175.53	59.66	98.98	-171.55	-91.30	-133.15	O1'-endo
A2	-67.47	171.78	58.85	130.45	-175.51	-86.70	-113.49	C2'-endo
G3	/	/	53.16	128.98	/	/	-112.36	C1'-exo
5'CTG								
C1	-66.55	175.02	58.05	106.07	-171.78	-91.18	-138.04	O1'-endo
T2	-72.99	173.58	54.56	121.25	-168.43	-88.47	-114.88	C1'-exo

(cont'd)

	α P-O5'	β O5'-C5'	γ C5'-C4'	δ C4'-C3'	ε C3'-O3'	ξ O3'-P	χ C1'-N	Conformation
G3	/	/	51.81	132.67	/	/	-102.15	C2'-endo
(GAG,CTC)								
5'GAG								
G1	-74.06	-178.12	62.59	105.19	-158.58	-83.05	-156.97	C1'-exo
A2	-69.78	171.86	49.71	130.83	-172.49	-90.48	-110.81	C2'-endo
G3	/	/	50.87	122.40	/	/	-113.08	C1'-exo
5'CTC								
C1	-66.97	175.39	56.96	119.66	-170.35	-87.19	-129.04	C1'-exo
T2	-66.97	169.73	51.77	124.35	-175.36	-89.06	-112.53	C1'-exo
C3	/	/	52.75	113.77	/	/	-118.42	C1'-exo
(TAG,ATC)								
5'TAG								
T1	-74.24	178.23	55.46	112.93	-164.57	-85.45	-135.21	C1'-exo
A2	-70.96	172.89	48.86	128.25	-173.86	-91.49	-110.01	C1'-exo
G3	/	/	50.17	117.31	/	/	-112.88	C1'-exo
5'CTA								
C1	-65.88	175.95	57.38	99.44	-171.04	-83.51	-146.38	O1'-endo
T2	-70.62	173.77	58.82	124.44	-169.25	-85.10	-122.12	C1'-exo
A3	/	/	49.78	129.88	/	/	-107.99	C1'-exo
(AXA,TX'T)								
(AAA,TTT)								
5'AAA								
A1	-70.84	174.29	59.24	83.72	-165.28	-73.14	-149.64	C3'-endo
A2	-66.33	176.50	56.78	85.13	-164.92	-70.52	-145.23	C3'-endo
A3	/	/	62.22	124.02	/	/	-123.09	C2'-endo
5'TTT								
T1	-69.21	171.89	58.07	121.84	-169.22	-87.03	-134.89	C1'-exo
T2	-68.06	170.76	55.14	117.39	-175.32	-91.05	-119.24	C1'-exo
T3	/	/	54.87	105.55	/	/	-128.90	O1'-endo

(cont'd)

	α P-O5'	β O5'-C5'	γ C5'-C4'	δ C4'-C3'	ε C3'-O3'	ξ O3'-P	χ C1'-N	Conformation
(AGA,TCT)								
5'AGA								
A1	-64.58	177.60	55.13	86.11	-159.11	-75.15	-160.16	C3'-endo
G2	-71.39	176.56	62.77	142.91	-171.71	-86.85	-123.60	C2'-endo
A3	/	/	48.78	115.51	/	/	-126.03	C1'-exo
5'TCT								
T1	-67.01	174.08	56.29	122.27	-173.35	-91.23	-130.08	C1'-exo
C2	-65.29	173.79	54.34	121.17	-174.15	-87.86	-123.93	C1'-exo
T3	/	/	54.90	123.05	/	/	-124.85	C1'-exo
(XGGX,X'CCX')								
(AGGA,TCCT)								
5'AGGA								
A1	-68.35	174.30	-173.13	137.22	-166.62	-113.80	-110.76	C2'-endo
G2	-68.94	171.56	52.15	136.82	-175.37	-90.73	-110.36	C2'-endo
G3	-68.81	174.72	54.34	107.89	-166.83	-88.20	-129.36	C1'-exo
A4	/	/	51.64	134.44	/	/	-102.93	C2'-endo
5'TCCT								
T1	/	/	/	/	/	-128.16	-116.59	/
C2	-69.65	169.65	/	135.27	-173.56	-92.10	-108.71	C2'-endo
C3	-66.82	171.05	54.61	104.22	-174.49	-88.94	-131.70	O1'-endo
T4	/	/	57.62	121.31	/	/	-120.20	C1'-exo
(CGGC,GCCG)								
5'CGGC								
C1	-77.51	179.38	58.83	112.52	-162.86	-89.96	-135.04	C1'-exo
G2	-65.54	168.66	50.45	130.46	-177.40	-90.43	-104.64	C2'-endo
G3	-66.68	168.94	53.56	108.36	-170.69	-96.53	-123.52	C1'-exo
C4	/	/	54.48	120.85	/	/	-120.53	C1'-exo
5'GCCG								
G1	-65.77	169.76	58.10	117.02	-176.60	-91.62	-120.38	C1'-exo
C2	-65.65	171.37	56.24	99.60	-175.42	-90.43	-134.15	O1'-endo
C3	-72.08	171.75	60.26	114.90	-166.22	-82.79	-125.63	C1'-exo

(cont'd)

	α P-O5'	β O5'-C5'	γ C5'-C4'	δ C4'-C3'	ε C3'-O3'	ξ O3'-P	χ C1'-N	Conformation
G4	/	/	50.65	127.39	/	/	-106.72	C1'-exo
(GGGG,CCCC)								
5'GGGG								
G1	-65.87	173.70	56.96	124.95	-172.52	-88.43	-133.17	C1'-exo
G2	-72.55	178.66	52.06	120.87	-170.96	-92.04	-129.91	C1'-exo
G3	-72.41	173.11	49.76	116.53	-168.72	-84.11	-130.72	C1'-exo
G4	/	/	49.30	116.97	/	/	-119.70	C1'-exo
5'CCCC								
C1	-66.82	174.64	58.57	101.08	-172.43	-86.65	-135.83	O1'-endo
C2	-66.88	170.65	54.84	125.90	-174.35	-89.82	-118.00	C1'-exo
C3	-70.68	179.56	55.51	112.96	-172.15	-91.19	-133.74	C1'-exo
C4	/	/	50.49	118.16	/	/	-133.14	C1'-exo
(TGGT,ACCA)								
5'TGGT								
T1	-71.30	169.44	61.56	109.09	-167.50	-85.34	-114.67	C1'-exo
G2	-79.11	148.08	48.61	140.05	-84.73	147.91	-91.82	C2'-endo
G3	-68.92	173.93	43.04	138.92	-174.31	-92.39	-118.79	C2'-endo
T4	/	/	54.15	119.21	/	/	-122.08	C1'-exo
5'ACCA								
A1	-65.33	170.72	58.50	125.70	-176.98	-98.99	-106.99	C1'-exo
C2	-63.07	172.67	55.99	115.67	-175.98	-91.23	-125.77	C1'-exo
C3	-80.46	173.61	59.62	125.36	-162.65	-85.53	-122.83	C1'-exo
A4	/	/	45.89	130.04	/	/	-94.48	C1'-exo

APPENDIX II

BACKBONE TORSION ANGLES OF THE EXPERIMENTAL SEQUENCES FROM NUCLEIC ACID DATABASE (NDB)

(a) 5'-G(1)

	α	β	γ	δ	ε	ξ	χ
Average value							
overall	-109.40	183.32	116.26	135.10	196.49	-114.87	-117.33
	(250.60)	(183.32)	(116.26)	(135.10)	(196.49)	(245.13)	(242.67)
NMR	-69.36	172.04	65.62	129.91	172.21	-102.78	-110.69
	(290.64)	(172.04)	(65.62)	(129.91)	(172.21)	(257.22)	(249.31)
X-Ray	-114.98	184.81	123.42	135.83	199.92	-116.58	-118.26
	(245.02)	(184.81)	(123.42)	(135.83)	(199.92)	(243.42)	(241.74)
Average value with terminated GGG only							
overall	n.a.	n.a.	238.28	138.38	187.01	-103.81	-118.61
	n.a.	n.a.	(238.28)	(138.38)	(187.01)	(256.19)	(241.39)
NMR	n.a.	n.a.	88.05	127.25	143.82	-65.50	-116.02
	n.a.	n.a.	(88.05)	(127.25)	(143.82)	(294.50)	(243.98)
X-Ray	n.a.	n.a.	239.82	124.67	195.26	-108.30	-120.78
	n.a.	n.a.	(239.82)	(124.67)	(195.26)	(251.70)	(239.22)
Average value with GGGG sequence							
overall	-149.57	192.18	170.53	134.56	192.41	-106.92	-121.35
	(210.43)	(192.18)	(170.53)	(134.56)	(192.41)	(253.08)	(238.65)
NMR	-57.06	161.13	64.10	136.80	184.65	-103.54	-124.12
	(302.94)	(161.13)	(64.10)	(136.80)	(184.65)	(256.46)	(235.88)
X-Ray	-158.52	195.19	178.51	134.39	192.99	-107.18	-121.14
	(201.48)	(195.19)	(178.51)	(134.39)	(192.99)	(252.83)	(238.86)
Standard deviation							
overall	94.82	34.91	102.86	19.60	35.83	46.91	34.78
NMR	16.39	15.85	34.07	14.42	25.34	33.61	20.35
X-Ray	99.84	36.50	107.03	20.26	35.73	48.38	36.34

(cont'd)

	α	β	γ	δ	ε	ξ	χ
Standard deviation with terminated GGG only							
overall	n.a.	n.a.	107.48	14.61	33.65	38.45	21.43
NMR	n.a.	n.a.	80.56	19.68	49.30	37.44	28.15
X-Ray	n.a.	n.a.	87.39	13.23	24.33	35.20	21.07
Standard deviation with GGGG sequence							
overall	120.74	43.50	123.31	24.06	33.25	46.35	23.22
NMR	3.70	3.10	1.58	13.46	5.40	15.87	4.79
X-Ray	122.87	44.45	124.25	24.78	34.41	47.95	24.06
NMR							
169D	n.a.	n.a.	180.63	108.04	178.35	-87.69	-130.51
1G5K	-65.86	192.16	64.59	143.60	173.80	-100.21	-122.14
1GJ1	-73.96	182.02	55.03	132.60	183.63	-118.70	-103.11
1IV6	-57.20	165.31	60.18	141.76	185.09	-97.26	-128.20
	-79.09	174.49	53.80	111.01	183.18	-96.98	-118.04
1J5N	-45.71	143.08	56.09	113.65	169.24	-97.84	-111.86
	n.a.	n.a.	49.62	126.36	87.36	-22.26	-83.58
1KBD	-59.46	159.36	62.38	143.94	189.97	-109.98	-124.00
1KVH	-83.33	198.38	60.88	110.99	170.85	-98.63	-89.47
1LFU	n.a.	n.a.	33.90	147.36	165.76	-86.54	-133.98
1LWA	-52.80	164.72	64.45	121.28	179.19	-85.46	-128.97
1N0K	-96.75	175.74	56.58	135.75	179.17	-165.65	-75.47
1N0O	-89.92	177.83	55.05	137.20	180.53	-156.49	-80.94
2KBD	-58.90	159.32	65.47	145.19	184.78	-115.19	-119.40
X-Ray							
PD0005	-59.21	238.19	27.59	150.67	181.60	-129.13	-95.31
	-233.71	162.85	176.82	147.11	196.57	-113.11	-140.19
PD0010	-58.18	167.90	55.93	147.17	212.40	-138.27	-92.16
	n.a.	197.13	59.67	144.09	199.05	-145.72	-85.85
PD0023	-119.81	213.69	71.95	108.57	153.60	-71.94	-117.33
	-159.89	244.85	105.47	91.18	168.11	-77.85	-120.92
PD0024	n.a.	n.a.	194.55	147.14	240.69	-171.67	-97.31
	n.a.	n.a.	187.86	146.19	240.63	-175.56	-97.95

(cont'd)

	α	β	γ	δ	ε	ξ	χ
PD0030	n.a.	n.a.	293.41	147.13	162.08	-149.03	-156.29
PD0032	-98.54	140.08	184.20	85.10	215.95	-81.57	-179.57
PD0033	n.a.	n.a.	3.79	152.19	163.01	-148.11	-165.20
PD0041	-90.26	179.66	58.81	147.81	209.76	-152.63	-58.40
PD0068	-163.05	229.09	63.44	149.38	222.40	-126.08	-106.06
PD0095	-279.17	190.08	181.67	147.04	189.12	-119.38	-118.60
	-271.38	193.84	176.70	146.40	189.25	-121.33	-121.38
PD0105	-329.64	209.08	279.95	153.36	170.98	-89.64	-124.66
PD0112	n.a.	n.a.	326.77	148.48	184.81	-88.14	-125.20
	-2.87	267.42	264.78	154.76	186.60	-74.50	-133.15
	n.a.	n.a.	338.76	148.67	182.98	-93.06	-123.28
	-349.02	268.51	271.14	149.82	184.64	-80.08	-135.72
	n.a.	n.a.	333.61	145.57	186.79	-90.59	-126.29
	-71.39	192.66	36.63	137.85	185.83	-101.37	-109.85
	n.a.	n.a.	332.10	147.48	188.58	-92.91	-121.84
	-355.43	262.67	269.96	150.18	183.09	-76.50	-130.67
	n.a.	n.a.	333.67	145.65	185.08	-88.57	-124.44
	-348.03	257.87	274.41	148.74	183.72	-78.08	-134.01
PD0145	-21.05	180.83	359.81	160.08	194.30	-110.39	-105.54
	-289.55	107.86	325.64	152.89	174.72	-127.30	-71.63
PD0146	-348.31	195.24	305.43	165.31	189.93	-93.39	-111.10
	-334.21	189.44	303.02	160.56	156.30	-72.81	-106.05
PD0155	-62.86	179.63	51.52	139.75	202.80	-137.78	-290.63
	-67.78	175.36	60.32	134.27	205.52	-127.00	-298.32
PD0165	-50.19	148.03	45.44	125.26	189.06	-113.96	-130.11
	-57.05	141.48	38.97	118.09	191.90	-113.74	-133.31
	-57.67	134.92	39.57	119.17	192.73	-115.13	-132.76
	-59.78	140.68	50.41	124.65	191.45	-117.16	-134.53
	-71.19	178.05	50.42	143.57	194.06	-151.05	-102.96
	-43.32	142.28	38.10	130.65	182.26	-106.93	-127.33
PD0192	-42.76	169.90	50.93	146.42	209.31	-102.08	-96.92
PD0200	-60.16	175.44	34.79	153.57	237.41	-178.97	-65.51
PD0219	-69.63	174.23	33.05	142.83	226.71	-145.18	-90.27
	-74.25	176.40	35.99	143.23	243.41	-172.24	-90.89
	-76.99	176.62	38.92	142.25	243.22	-170.68	-91.91

(cont'd)

	α	β	γ	δ	ε	ξ	χ
	-72.47	175.66	35.58	141.27	226.02	-144.69	-91.01
PD0221	-205.05	218.81	164.15	136.69	244.87	-163.59	-112.24
PD0222	n.a.	n.a.	298.95	150.49	236.57	-76.71	-81.76
PD0230	-43.14	142.11	41.06	133.80	181.37	-104.09	-127.16
PD0231	-46.01	141.56	44.34	133.48	181.78	-102.92	-123.08
PD0251	-59.86	151.23	36.17	145.60	195.28	-94.67	-115.89
PD0252	-262.97	216.93	190.27	143.56	208.24	-74.46	-143.89
PD0259	-78.46	174.03	44.85	150.85	309.17	107.94	-78.00
	-56.49	163.15	48.83	139.85	193.71	-139.56	-95.70
PD0270	-282.06	226.18	179.06	101.39	221.64	-96.27	-171.48
PD0286	-51.70	179.30	41.37	136.15	272.40	-211.83	-84.49
	-64.37	201.33	29.31	139.27	235.00	-167.25	-89.56
	-56.03	174.06	50.12	138.74	170.62	-121.66	-106.23
	-29.78	179.73	52.36	142.56	181.53	-103.89	-101.74
PD0307	-209.67	190.89	179.60	137.91	224.96	-83.84	-154.07
	61.94	204.88	172.70	89.81	170.72	-48.88	-139.50
PD0309	-96.61	214.70	50.03	137.46	136.39	-75.84	-79.96
	-76.35	154.43	56.40	145.03	204.24	-104.30	-78.57
PD0310	-69.93	183.56	30.65	139.32	284.16	-201.38	-98.71
PD0312	-51.21	142.90	42.02	130.47	182.42	-106.37	-133.92
	-50.57	143.22	47.97	137.52	177.10	-99.41	-116.38
PD0316	-67.08	170.57	53.70	147.63	239.81	-190.08	-116.61
	-230.77	150.61	170.79	146.49	184.32	-106.24	-135.85
PD0319	-62.42	191.26	48.06	148.66	170.84	-124.58	-76.03
PD0320	-49.71	187.77	38.62	145.82	151.30	-101.30	-84.67
PD0338	-92.99	192.35	55.73	161.69	319.34	-236.91	-77.07
PD0364	-54.53	184.63	43.16	142.35	182.98	-135.16	-93.02
PD0365	-59.16	184.13	47.88	142.80	193.79	-137.48	-88.97
PD0393	n.a.	n.a.	333.27	108.44	175.84	-55.59	-133.48
	-103.60	174.13	84.89	109.08	196.13	-96.70	-125.26
	-62.53	178.98	43.68	117.68	58.26	-308.11	-98.70
	-167.94	206.79	60.20	151.50	314.87	-217.40	-84.67
	-81.88	147.64	13.14	123.72	156.71	-90.34	-100.82
PD0424	-58.75	147.27	39.98	141.38	204.99	-115.61	-131.97
	-72.70	124.38	54.34	106.79	190.80	-124.55	-141.06

(cont'd)

	α	β	γ	δ	ε	ξ	χ
PD0429	-8.69	293.87	255.56	145.77	171.86	-85.70	-121.14
	-76.30	208.90	48.20	140.95	163.33	-86.02	-92.97
	-295.22	266.90	201.65	141.49	226.65	-153.99	-124.21
	-42.11	138.18	45.46	137.73	179.17	-126.59	-106.63
PD0504	n.a.	n.a.	289.55	136.23	191.56	-96.56	-106.42
	-52.70	172.01	35.42	139.54	225.51	-178.82	-94.27
	-50.27	141.45	43.44	138.87	191.73	-93.26	-113.56
PD0505	n.a.	n.a.	281.74	127.08	200.77	-98.29	-113.26
	-56.19	163.35	45.42	140.02	231.74	-177.30	-94.09
	n.a.	n.a.	280.16	135.92	194.23	-91.44	-118.25
	-64.54	172.88	46.97	143.12	231.25	-171.17	-100.40
PDE003	n.a.	n.a.	220.27	113.25	192.07	-122.19	-109.36
	n.a.	n.a.	214.75	147.53	196.01	-137.40	-105.12
PDE011	n.a.	220.21	105.61	125.36	195.88	-84.19	-119.98
PDE012	n.a.	214.59	69.57	145.30	219.99	-107.85	-110.70
	n.a.	244.04	37.48	163.60	171.43	-89.94	-92.06
PDE0128	-145.06	189.81	73.06	82.01	207.23	-70.05	-172.87
	-79.51	181.57	49.90	90.56	210.60	-66.96	-162.61
	-78.19	169.78	57.67	73.49	196.67	-64.32	-172.92
	-92.86	182.51	71.48	76.53	219.43	-93.26	-176.20
	-300.37	158.36	55.71	67.74	204.36	-61.52	-168.17

(b) 5'-G(2)

	α	β	γ	δ	ε	ξ	χ
Average value							
overall	-90.62	-177.62	68.37	130.53	191.46	-106.80	-116.98
	(269.38)	(182.38)	(68.37)	(130.53)	(191.46)	(253.20)	(243.02)
NMR	-90.09	-166.41	76.60	133.50	185.17	-107.94	-118.12
	(269.91)	(193.59)	(76.60)	(133.50)	(185.17)	(252.06)	(241.88)

(cont'd)

	α	β	γ	δ	ε	ξ	χ
X-Ray	-90.68	-179.09	67.29	130.14	192.29	-106.66	-116.83
	(269.32)	(180.91)	(67.29)	(130.14)	(192.29)	(253.34)	(243.17)
Average value with terminated GGG only							
overall	-68.00	-191.93	51.30	128.49	195.60	-106.31	-116.01
	(292.00)	(168.07)	(51.30)	(128.49)	(195.60)	(253.69)	(243.99)
NMR	-80.45	-172.24	57.71	130.46	179.64	-104.55	-118.55
	(279.55)	(187.76)	(57.71)	(130.46)	(179.64)	(255.45)	(241.45)
X-Ray	-73.18	-192.63	53.72	127.75	201.77	-108.10	-118.02
	(286.83)	(167.37)	(53.72)	(127.75)	(201.77)	(251.90)	(241.98)
Average value with GGGG sequence							
overall	-79.92	-174.03	53.06	128.78	194.66	-108.78	-117.43
	(280.08)	(185.97)	(53.06)	(128.78)	(194.66)	(251.22)	(242.57)
NMR	-74.92	-175.51	61.26	135.55	182.47	-104.93	-121.16
	(285.08)	(184.49)	(61.26)	(135.55)	(182.47)	(255.07)	(238.84)
X-Ray	-80.29	-173.92	52.45	128.27	195.57	-109.06	-117.14
	(279.71)	(186.08)	(52.45)	(128.27)	(195.57)	(250.94)	(242.86)
Standard deviation							
overall	71.54	56.86	61.21	22.89	31.45	43.09	26.20
NMR	61.08	31.36	44.12	9.90	16.84	18.90	13.44
X-Ray	73.08	59.35	63.21	24.08	32.86	45.38	27.49
Standard deviation with terminated GGG only							
overall	25.29	31.29	27.19	18.35	20.92	35.35	17.90
NMR	20.78	16.65	2.70	9.92	7.21	13.50	8.93
X-Ray	25.94	32.34	29.60	19.75	21.40	38.40	19.29
Standard deviation with GGGG sequence							
overall	48.33	59.41	47.21	22.90	44.55	56.91	30.93
NMR	25.43	20.18	6.82	11.96	8.80	13.80	11.56
X-Ray	49.80	61.48	48.91	23.53	46.06	58.96	32.00

(cont'd)

	α	β	γ	δ	ε	ξ	χ
NMR							
169D	-73.01	-182.22	57.83	128.27	179.81	-111.45	-122.86
1G5K	-80.26	-162.15	57.93	142.10	187.14	-105.28	-124.73
1GJ1	-60.54	-186.84	66.36	138.31	167.14	-92.50	-93.81
1IV6	-67.39	-185.30	59.04	129.96	182.79	-109.75	-124.14
	-65.30	-193.63	61.39	118.43	180.81	-93.71	-126.49
1J5N	-24.38	-182.88	28.10	120.98	173.20	-90.06	-112.02
1KBD	-64.41	-181.49	54.95	141.29	186.75	-113.22	-124.51
1KVH	-67.27	-177.39	61.89	123.91	190.91	-86.43	-97.96
1LFU	-64.78	-162.64	73.51	146.72	162.26	-111.50	-103.67
1LWA	-103.93	-153.02	60.34	121.81	172.35	-88.99	-108.29
1N0K	-233.07	-112.18	178.42	141.51	218.93	-148.13	-132.33
1N0O	-210.38	-91.52	167.52	138.64	216.85	-139.62	-134.04
2KBD	-56.43	-192.02	68.49	143.56	188.30	-112.56	-130.68
X-Ray							
PD0005	-252.28	-118.45	174.09	152.00	182.40	-54.16	-155.37
	-326.08	-232.57	318.49	151.32	215.00	-137.26	-108.36
PD0010	-54.49	-218.33	53.45	116.66	180.49	-88.92	-114.99
	-36.60	-218.04	47.78	137.60	196.10	-118.11	-106.29
PD0023	-103.92	-161.16	75.61	111.30	185.71	-75.29	-128.80
	-77.07	175.15	67.20	92.76	163.13	-37.78	-137.99
PD0024	-72.75	-210.92	46.49	139.87	183.03	-96.46	-116.03
	-75.18	-207.90	45.26	140.84	190.02	-101.16	-117.31
PD0030	-84.30	-185.68	55.67	84.21	196.57	-57.84	-159.22
PD0032	-75.04	-186.62	60.80	85.85	196.76	-57.05	-162.29
PD0033	-76.67	-191.82	50.37	82.22	187.34	-43.50	-156.52
PD0041	-60.21	-223.94	62.17	149.08	181.72	-121.94	-98.77
PD0068	-61.16	162.62	29.99	145.04	272.59	85.87	-244.42
PD0095	-51.30	-170.80	42.19	135.22	273.42	-157.47	-94.42
	-58.15	-168.84	43.54	130.43	271.49	-159.16	-93.47
PD0105	-68.90	-145.06	40.00	154.15	182.73	-116.08	-99.20
PD0112	-73.79	-165.79	42.20	130.86	188.16	-92.88	-114.06
	-81.72	-166.10	43.70	137.11	191.87	-109.70	-111.60
	-62.16	-168.48	35.67	132.85	185.07	-94.57	-111.70

(cont'd)

	α	β	γ	δ	ε	ξ	χ
	-85.07	-156.28	44.03	139.57	184.18	-102.26	-104.99
	-63.92	-170.10	38.29	130.61	191.63	-94.42	-116.04
	-71.39	-167.34	36.63	137.85	185.83	-101.37	-109.85
	-60.91	-174.20	36.27	130.34	185.76	-92.04	-111.71
	-80.83	-159.41	41.82	135.83	182.23	-99.06	-106.79
	-68.41	-167.86	41.00	134.38	190.37	-98.39	-113.15
	-82.70	-159.71	44.53	141.69	189.57	-108.02	-109.75
PD0145	-45.45	-158.16	-1.11	139.20	113.44	-57.36	-102.32
	-209.42	-307.94	133.59	176.93	251.65	-108.64	-102.32
PD0146	-35.11	-153.06	-16.71	138.48	119.95	-56.75	-107.42
	-104.48	-146.37	53.18	101.02	114.59	-57.80	-113.99
PD0155	-63.05	-195.87	56.39	143.72	179.64	-93.65	-107.78
	-67.69	-190.98	55.68	147.72	182.06	-100.63	-101.84
PD0165	-60.28	-176.04	54.29	142.58	178.36	-85.21	-105.01
	-54.86	-191.77	60.12	143.72	197.50	-114.37	-117.92
	-51.78	-185.05	52.73	142.53	183.57	-89.60	-108.34
	-59.51	-188.76	60.28	139.64	176.56	-81.53	-123.21
	-53.49	-210.91	62.92	140.85	182.48	-86.41	-130.13
	-61.79	-162.33	48.66	142.38	172.04	-83.48	-107.12
PD0192	-55.05	-202.57	36.43	142.11	252.27	-203.46	-75.80
PD0200	-67.26	-222.07	49.35	147.54	210.76	-150.80	-89.90
PD0219	-85.69	-228.59	63.74	70.74	189.90	-69.03	-160.44
	-307.19	-134.07	202.46	85.29	213.82	-66.69	-169.92
	-307.39	-132.27	203.94	85.32	214.06	-65.68	-168.69
	-82.20	-229.00	61.26	71.69	185.43	-65.82	-161.92
PD0221	-44.60	-227.42	31.66	136.01	171.11	-106.62	-92.35
PD0222	-109.33	-296.39	148.21	150.57	242.20	-118.16	-122.57
PD0230	-58.62	-172.54	49.76	144.23	198.03	-123.48	-120.73
PD0231	-58.53	-171.07	45.54	141.59	175.62	-95.56	-115.22
PD0251	-74.61	-189.36	52.63	135.95	195.79	-135.47	-103.68
PD0252	-72.72	-187.24	45.04	140.14	210.22	-149.12	-97.96
PD0259	-180.32	-140.64	75.60	131.03	195.80	-122.88	-112.59
	-41.38	-200.70	35.87	134.98	202.42	-120.16	-113.60
PD0270	-29.21	-197.90	44.06	135.11	187.45	-95.12	-131.05
PD0286	-77.05	-218.11	44.35	145.00	182.59	-115.77	-97.17

(cont'd)

	α	β	γ	δ	ε	ξ	χ
	-49.60	-230.05	42.05	140.85	180.29	-143.83	-97.87
	-55.52	-178.91	49.64	145.30	193.82	-158.78	-84.25
	-42.39	-181.46	47.88	142.21	177.26	-78.03	-115.06
PD0307	-68.48	-196.88	37.84	140.71	250.38	-190.95	-88.57
	-86.05	-171.24	52.15	80.46	144.98	-86.68	-134.41
PD0309	-56.02	-141.61	39.09	148.61	200.84	-181.24	-95.29
	-89.33	-197.77	56.81	141.38	185.04	-140.46	-80.93
PD0310	-297.32	-133.72	196.00	141.33	219.41	-135.50	-131.15
PD0312	-46.85	-176.07	39.60	140.74	181.58	-106.36	-122.53
	-72.63	-164.43	52.42	141.39	199.44	-104.88	-106.71
PD0316	-264.06	-105.71	177.90	135.65	207.48	-88.60	-151.97
	-316.93	-220.40	312.08	147.67	180.72	-113.40	-105.18
PD0319	-52.79	-185.01	56.44	149.79	162.29	-93.04	-99.58
PD0320	-46.11	-164.63	52.29	155.89	190.51	-128.36	-90.46
PD0338	-149.91	-165.56	57.01	143.88	203.85	-121.76	-103.35
PD0364	-41.66	-193.31	40.91	144.14	188.94	-138.57	-104.52
PD0365	-41.02	-198.22	39.29	141.88	179.72	-145.38	-90.29
PD0393	-103.60	-185.87	84.89	109.08	196.13	-96.70	-125.26
	-62.53	-181.02	43.68	117.68	58.26	-308.11	-98.70
	-167.94	-153.21	60.20	151.50	314.87	-217.40	-84.67
	-81.88	-212.36	13.14	123.72	156.71	-90.34	-100.82
	-84.44	-164.96	65.94	107.13	160.95	-56.78	-123.53
PD0424	-55.91	-186.04	41.91	140.44	190.81	-116.46	-130.09
	-318.41	-192.66	315.33	170.93	187.40	-100.84	-124.34
PD0429	-76.30	-151.10	48.20	140.95	163.33	-86.02	-92.97
	-49.57	-167.74	40.44	137.75	189.12	-118.21	-100.50
	-42.11	-221.82	45.46	137.73	179.17	-126.59	-106.63
	-319.53	-209.10	321.03	151.35	189.73	-77.35	-91.35
PD0504	-52.70	-187.99	35.42	139.54	225.51	-178.82	-94.27
	-62.05	-179.90	36.44	140.66	199.53	-139.78	-93.56
	-40.42	-211.16	38.00	138.35	206.92	-164.87	-107.15
PD0505	-56.19	-196.65	45.42	140.02	231.74	-177.30	-94.09
	-55.67	-214.24	39.85	137.59	199.79	-98.62	-113.07
	-64.54	-187.12	46.97	143.12	231.25	-171.17	-100.40
	-67.68	-211.86	49.62	139.65	186.90	-91.02	-117.14

PDE003	-23.87	-201.98	21.80	136.18	181.93	-100.91	-92.64
	-2.24	-231.14	27.60	125.24	170.80	-91.91	-103.53
PDE011	-84.99	-184.10	51.23	119.27	164.51	-94.36	-101.99
PDE012	-69.61	-229.23	79.04	75.49	195.32	-74.14	-153.71
	-94.64	-142.32	37.16	141.64	161.94	-102.30	-78.09
PDE0128	-79.51	-178.43	49.90	90.56	210.60	-66.96	-162.61
	-78.19	-190.22	57.67	73.49	196.67	-64.32	-172.92
	-92.86	-177.49	71.48	76.53	219.43	-93.26	-176.20
	-59.63	-201.64	55.71	67.74	204.36	-61.52	-168.17
	-70.99	-198.60	61.98	73.15	204.04	-88.80	-177.18

(c) 5'-G(3)

	α	β	γ	δ	ε	ξ	χ
Average value							
overall	-85.98	-185.20	62.97	129.59	191.31	-113.09	-116.29
	(274.02)	(174.80)	(62.97)	(129.59)	(191.31)	(246.91)	(243.71)
NMR	-74.42	-187.14	58.03	130.12	162.66	-125.86	-118.40
	(285.58)	(172.86)	(58.03)	(130.12)	(162.66)	(234.14)	(241.60)
X-Ray	-87.50	-184.94	63.62	129.52	193.50	-112.11	-116.01
	(272.50)	(175.06)	(63.62)	(129.52)	(193.50)	(247.89)	(243.99)
Average value with GGGG sequence							
overall	-87.60	-185.87	64.25	126.73	191.55	-111.98	-117.34
	(272.40)	(174.13)	(64.25)	(126.73)	(191.55)	(248.02)	(242.66)
NMR	-64.45	-175.11	56.94	133.05	173.55	-104.42	-121.37
	(295.55)	(184.89)	(56.94)	(133.05)	(173.55)	(255.58)	(238.63)
X-Ray	-89.34	-186.68	64.79	126.25	193.10	-112.63	-117.04
	(270.66)	(173.32)	(64.79)	(126.25)	(193.10)	(247.37)	(242.96)
Standard deviation							
overall	65.85	28.03	52.65	21.74	36.35	52.25	22.01
NMR	11.52	36.77	8.11	16.80	79.97	64.09	14.65
X-Ray	69.82	26.90	55.93	22.37	30.55	51.53	22.84

(cont'd)

	α	β	γ	δ	ε	ξ	χ
Standard deviation with GGGG sequence							
overall	61.92	29.33	52.45	21.92	34.79	68.88	24.61
NMR	0.69	3.33	5.67	18.08	22.60	8.85	4.10
X-Ray	63.91	30.27	54.37	22.30	35.44	71.78	25.50
NMR							
169D	-78.07	-176.59	63.55	130.69	179.76	-86.67	-129.76
1G5K	-82.85	-174.78	48.25	87.03	n.a.	n.a.	-134.06
1GJ1	-72.17	-162.88	48.56	147.26	n.a.	n.a.	-95.84
1IV6	-66.57	-271.83	59.20	134.39	288.91	-101.98	-122.00
	-67.29	-240.70	57.84	130.30	n.a.	n.a.	-123.47
1J5N	-82.37	-164.36	46.16	130.96	23.32	-269.85	-100.82
1KBD	-63.72	-173.40	52.36	142.47	187.18	-108.98	-118.80
1KVH	-105.07	-134.44	57.43	115.27	n.a.	n.a.	-97.23
1LFU	-80.71	-163.04	70.94	146.50	126.00	-109.29	-103.68
1LWA	-65.08	-178.95	63.29	112.21	147.46	-94.22	-126.10
1N0K	-68.88	-208.61	59.68	134.04	n.a.	n.a.	-142.02
1N0O	-70.14	-210.31	71.94	135.98	n.a.	n.a.	-126.22
2KBD	-64.55	-172.99	55.19	144.49	186.00	-110.06	-119.22
X-Ray							
PD0005	-224.31	-223.48	178.83	149.72	164.91	-110.82	-116.41
	-93.62	-208.68	63.08	107.37	195.86	-106.46	-159.41
PD0010	-51.81	-186.86	48.37	144.21	255.55	-171.38	-98.56
	-41.94	-203.85	43.09	141.53	261.24	-176.02	-99.06
PD0023	-89.32	-186.28	61.04	101.98	154.29	-95.01	-133.61
	-115.32	-158.12	49.94	87.61	137.58	-78.02	-128.57
PD0024	-63.50	-170.65	38.33	139.60	183.97	-117.25	-104.18
	-60.88	-176.40	40.76	133.61	185.82	-104.61	-108.17
PD0030	-76.37	-168.02	49.43	81.91	187.94	-75.36	-145.11
PD0032	-75.64	-168.43	49.78	80.16	188.71	-72.95	-148.43
PD0033	-85.33	-162.23	49.71	79.59	183.01	-71.48	-145.44
PD0041	-59.21	-193.36	61.75	138.18	232.85	-52.02	-117.99
PD0068	-82.52	-192.10	56.72	130.70	229.62	-81.57	-92.56
PD0095	-40.91	-214.85	48.68	138.86	170.95	-86.19	-94.29

(cont'd)

	α	β	γ	δ	ε	ξ	χ
	-59.56	-198.45	53.22	141.96	170.25	-82.05	-94.74
PD0105	37.66	-158.31	295.62	153.95	174.01	-84.25	-122.29
PD0112	-68.09	-181.09	48.18	140.41	204.01	-174.24	-102.57
	-50.09	-202.29	51.05	113.72	n.a.	n.a.	-136.51
	-57.59	-181.71	40.14	143.70	203.21	-167.57	-99.05
	-57.54	-191.71	48.54	120.08	n.a.	n.a.	-127.11
	-63.32	-185.39	44.17	142.86	207.04	-176.49	-101.43
	-57.47	-189.01	50.91	117.13	n.a.	n.a.	-126.43
	-69.92	-180.77	48.61	145.75	205.85	178.74	-97.73
	-54.02	-192.33	50.41	116.38	n.a.	n.a.	-129.26
	-65.93	-182.95	45.58	144.31	208.62	-178.59	-95.46
	-54.76	-194.21	48.41	123.20	n.a.	n.a.	-126.18
PD0145	-145.49	-102.11	74.49	115.65	173.04	-124.77	-99.75
	-71.69	-168.31	53.64	149.81	193.33	-118.17	-104.33
PD0146	-144.77	-110.53	72.86	107.96	164.32	-102.19	-111.91
	-147.50	-121.58	98.71	113.15	172.10	-116.40	-133.80
PD0155	-69.77	-172.96	42.57	141.22	202.83	-136.07	-101.70
	-70.40	-182.81	53.84	121.12	194.06	-130.01	-119.60
PD0165	-63.15	-174.24	41.26	129.13	189.39	-106.93	-105.68
	-330.94	-180.05	299.10	147.11	177.05	-83.24	-105.15
	-69.64	-173.14	40.48	125.10	181.64	-93.17	-102.21
	-83.03	-160.47	49.23	133.06	183.82	-101.78	-106.02
	-79.61	-159.20	44.96	136.01	189.36	-112.34	-108.64
	-64.10	-166.24	38.26	132.11	181.35	-95.30	-104.54
PD0192	-69.60	-224.20	48.26	141.38	178.63	-109.12	-111.01
PD0200	-71.50	-202.23	56.89	147.81	185.70	-92.24	-122.07
PD0219	-64.07	-181.08	53.36	85.84	192.50	-71.54	-160.85
	-54.09	-186.10	47.34	88.50	190.94	-72.67	-162.78
	-52.96	-184.28	42.97	88.06	186.55	-68.38	-160.58
	-71.67	-177.05	56.27	86.46	193.60	-73.89	-161.98
PD0221	-47.87	-176.05	38.20	138.91	207.53	-108.29	-116.65
PD0222	-141.12	-228.60	91.85	150.70	207.30	-150.25	-128.75
PD0230	-333.52	-183.63	-48.11	156.85	174.66	-84.65	-99.47
PD0231	-65.91	-161.57	40.80	139.19	177.56	-93.36	-102.72
PD0251	-66.19	-193.95	55.74	135.70	197.05	-123.92	-113.19

(cont'd)

	α	β	γ	δ	ε	ξ	χ
PD0252	-55.97	-207.22	45.71	143.99	194.35	-134.85	-103.20
PD0259	-45.11	-200.15	48.07	141.24	173.62	-95.60	-106.76
	-51.80	-197.96	46.24	137.45	175.37	-103.37	-100.86
PD0270	-41.26	-163.17	22.91	149.25	191.79	-139.44	-108.65
PD0286	-40.73	-197.44	49.09	140.94	192.57	-144.98	-94.35
	-37.66	-199.04	48.78	143.65	196.26	-121.32	-112.88
	-13.43	-242.35	46.20	140.35	204.76	-165.26	-105.16
	-67.57	-166.20	38.94	140.13	196.91	-145.82	-92.38
PD0307	-78.84	-223.40	49.52	140.14	207.99	-128.15	-126.63
	-180.56	-123.25	151.75	92.78	196.44	-60.77	-158.18
PD0309	-239.74	-123.21	181.32	129.10	206.84	-89.40	-151.36
	-49.90	-199.75	60.37	147.99	207.93	-120.36	-94.94
PD0310	-53.89	-204.75	45.13	140.16	195.72	-121.55	-105.94
PD0312	-47.96	-178.48	43.86	140.93	n.a.	n.a.	-114.14
	-49.77	-197.25	36.34	140.38	n.a.	n.a.	-87.86
PD0316	-228.61	-188.68	177.14	144.90	180.82	-111.86	-113.01
	-85.94	-145.52	44.25	158.17	147.67	-136.26	-111.77
PD0319	-73.77	-149.06	46.10	155.80	294.44	-233.06	-83.21
PD0320	-67.34	-195.60	60.23	167.82	270.54	-206.81	-91.48
PD0338	-45.98	-205.24	40.58	134.17	169.37	-98.56	-97.68
PD0364	-8.23	-221.14	25.98	129.74	166.63	-101.31	-111.72
PD0365	-333.21	-234.76	3.63	131.75	161.61	-100.10	-86.92
PD0393	-62.53	-181.02	43.68	117.68	58.26	-308.11	-98.70
	-167.94	-153.21	60.20	151.50	314.87	-217.40	-84.67
	-81.88	-212.36	13.14	123.72	156.71	-90.34	-100.82
	-84.44	-164.96	65.94	107.13	160.95	-56.78	-123.53
	-89.13	-170.25	53.20	103.53	170.27	-74.53	-135.09
PD0424	-48.37	-177.08	33.83	138.89	185.62	-99.05	-109.10
	-81.03	-155.83	44.37	145.27	188.66	-105.17	-94.62
PD0429	-49.57	-167.74	40.44	137.75	189.12	-118.21	-100.50
	-39.14	-200.09	39.69	138.73	205.76	-158.69	-108.47
	-319.53	-209.10	321.03	151.35	189.73	-77.35	-91.35
	-113.92	-279.15	166.67	146.66	201.71	-90.37	-146.72
PD0504	-50.27	-218.55	43.44	138.87	191.73	-93.26	-113.56
	-62.05	-179.90	36.44	140.66	199.53	-139.78	-93.56

(cont'd)

	α	β	γ	δ	ε	ξ	χ
	-40.42	-211.16	38.00	138.35	206.92	-164.87	-107.15
PD0505	-55.67	-214.24	39.85	137.59	199.79	-98.62	-113.07
	-50.94	-194.24	31.94	134.59	200.08	-155.28	-97.84
	-67.68	-211.86	49.62	139.65	186.90	-91.02	-117.14
	-68.89	-175.69	43.54	140.39	195.25	-128.71	-100.74
PDE003	-52.90	-191.87	58.45	138.32	221.72	-128.68	-113.80
	-67.38	-189.83	74.76	140.16	216.47	-122.86	-108.80
PDE011	-68.78	-166.80	56.05	151.15	206.28	-142.30	-107.35
PDE012	-69.91	-188.01	56.31	142.21	202.50	-127.09	-104.33
	-56.78	-177.64	48.89	141.39	184.44	-124.40	-102.10
PDE0128	-78.19	-190.22	57.67	73.49	196.67	-64.32	-172.92
	-92.86	-177.49	71.48	76.53	219.43	-93.26	-176.20
	-59.63	-201.64	55.71	67.74	204.36	-61.52	-168.17
	-289.01	-198.60	61.98	73.15	204.04	-88.80	-177.18
	-236.53	-165.14	222.32	125.66	212.55	-82.66	-164.76

(d) 5'-G(4)

	α	β	γ	δ	ε	ξ	χ
Average value							
overall	-86.67	177.79	60.82	124.16	196.03	-112.94	-119.54
	(273.33)	(177.79)	(60.82)	(124.16)	(196.03)	(247.06)	(240.46)
NMR	-64.45	226.89	56.94	133.06	179.55	-104.42	-121.37
	(295.55)	(226.89)	(56.94)	(133.06)	(179.55)	(255.58)	(238.63)
X-Ray	-89.05	172.53	61.24	123.21	198.17	-114.05	-119.35
	(270.95)	(172.53)	(61.24)	(123.21)	(198.17)	(245.95)	(240.65)
Standard deviation							
overall	56.61	38.89	38.44	21.64	31.03	37.06	25.37
NMR	0.69	69.43	5.67	18.08	12.21	8.85	4.10
X-Ray	59.16	32.05	40.46	22.05	32.23	39.28	26.71

(cont'd)

	α	β	γ	δ	ε	ξ	χ
NMR							
1KBD	-63.72	186.60	52.36	142.47	187.18	-108.98	-118.80
1LWA	-65.08	307.05	63.29	112.21	165.46	-94.22	-126.10
2KBD	-64.55	187.01	55.19	144.49	186.00	-110.06	-119.22
X-Ray							
PD0068	-82.52	167.90	56.72	130.70	229.62	-81.57	-92.56
PD0095	-40.91	145.15	48.68	138.86	170.95	-86.19	-94.29
	-59.56	161.55	53.22	141.96	170.25	-82.05	-94.74
PD0112	-50.09	157.71	51.05	113.72	n.a.	n.a.	-136.51
	-57.54	168.29	48.54	120.08	n.a.	n.a.	-127.11
	-57.47	170.99	50.91	117.13	n.a.	n.a.	-126.43
	-54.02	167.67	50.41	116.38	n.a.	n.a.	-129.26
	-54.76	165.79	48.41	123.20	n.a.	n.a.	-126.18
PD0145	-145.49	257.89	74.49	115.65	173.04	-124.77	-99.75
PD0146	-144.77	249.47	72.86	107.96	164.32	-102.19	-111.91
PD0251	-74.61	170.64	52.63	135.95	195.79	-135.47	-103.68
	-66.19	166.05	55.74	135.70	197.05	-123.92	-113.19
PD0252	-72.72	172.76	45.04	140.14	210.22	-149.12	-97.96
	-55.97	152.78	45.71	143.99	194.35	-134.85	-103.20
PD0393	-167.94	206.79	60.20	151.50	314.87	-217.40	-84.67
	-81.88	147.64	13.14	123.72	156.71	-90.34	-100.82
	-84.44	195.04	65.94	107.13	160.95	-56.78	-123.53
	-89.13	189.75	53.20	103.53	170.27	-74.53	-135.09
PD0429	-39.14	159.91	39.69	138.73	205.76	-158.69	-108.47
	-113.92	80.85	166.67	146.66	201.71	-90.37	-146.72
PD0504	-62.05	180.10	36.44	140.66	199.53	-139.78	-93.56
	-40.42	148.84	38.00	138.35	206.92	-164.87	-107.15
PD0505	-50.94	165.76	31.94	134.59	200.08	-155.28	-97.84
	-68.89	184.31	43.54	140.39	195.25	-128.71	-100.74
PDE0128	-92.86	182.51	71.48	76.53	219.43	-93.26	-176.20
	-59.63	158.36	55.71	67.74	204.36	-61.52	-168.17
	-289.01	161.40	61.98	73.15	204.04	-88.80	-177.18
	-236.53	194.86	222.32	125.66	212.55	-82.66	-164.76

REFERENCES

1. Dahm, R. *Dev. Biol.* **2005**, *278*, 274.
2. Levene, P. A. *J. Biol. Chem.* **1919**, *40*, 415.
3. Astbury, W. T. *Symp. Soc. Exp. Biol. (Nucleic Acids)*, **1947**, *1*, 66.
4. Watson, J.; Crick, F. *Nature* **1953**, *171*, 737.
5. Lehninger, A. L.; Nelson, D. L.; Cox, M. M. *Principles of Biochemistry* **1993**, 2*nd ed.*, Worth Publishers, Inc.
6. Belmont, P.; Constant, J. F.; Demeunynck, M. *Chem. Soc. Rev.* **2001**, *30*, 70.
7. Markley, J. L.; Bax, A.; Arata, Y,; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wuthrich, K. *Pure & Appl. Chem.* **1998**, 70, 117.
8. Dickerson, R. E.; Bansal, M.; Calladine, C. R.; Diekmann, S.; Hunter, W. N.; Kennard, O.; von Kitzing, E.; Lavery, R.; Nelson, H. C. M.; Olso, W. K.; Saenger, W.; Shakked, Z.; Sklenar, H.; Soumpasis, D. M.; Tung, C.-S.; Wang, H. J.; Zhurkin, V. B. *EMBO J.* **1989**, *8*, 1.
9. Saenger, W. *Principles of Nucleic Acid Structure* **1984**, Springer-Verlag.
10. Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247.
11. Prat, F.; Houk, K. N.; Foote, C. S. *J. Am. Chem. Soc.* **1998**, *120*, 845.
12. Kino, K.; Saito, I.; Sugiyama, H. *J. Am. Chem. Soc.* **1998**, *120*, 7373.
13. Weatherly, S. C.; Yang, I. V.; Thorp, H. H. *J. Am. Chem. Soc.* **2001**, *123*, 1236.
14. Giese, B. *Acc. Chem. Res.* **2000**, *33*, 631.
15. Sugiyama, H.; Saito, I. *J. Am. Chem. Soc.* **1996**, *118*, 7063.

16. Saito, I.; Takayama, M.; Sugiyama, H.; Nakatani, K. *J. Am. Chem. Soc.* **1995**, *117*, 6406.

17. Yoshioka, Y.; Kitagawa, Y.; Takano, Y.; Yamaguchi, K.; Nakamura, T.; Saito, I. *J. Am. Chem. Soc.* **1999**, *121*, 8712.

18. Fu, A.Y.Y. *M. Phil. Thesis*, The Chinese University of Hong Kong, **2004**.

19. Go, R. S.; Adjei, A. A. *J.Clin. Oncol.* **1999**, *17*, 409.

20. Rosenberg, B.; van Camp, L.; Trosko, J. L.; Mansour, V. H. *Nature* **1969**, *222*, 385.

21. Matsui, T.; Shigeta, Y.; Hirao, K. *J. Phys. Chem. B* **2007**, *111*, 1176.

22. Jamieson, E. R.; Lippard, S. J. *Chem. Rev.* **1999**, *99*, 2467.

23. Fuertes, M. A.; Alonso, C.; Perez, J. M. *Chem. Rev.* **2003**, *103*, 645.

24. Baik, M.; Friesner, R. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2003**, *125*, 14082.

25. Spiegel, K.; Rothlisbergeer, U.; Carloni, P. *J. Phys. Chem. B* **2004**, *108*, 2699.

26. Yang, X. L.; Wang, A. H.-J. *J. Pharmacol. Ther.* **1999**, *83*, 181.

27. Eastman, A. *Biochemistry* **1986**, *25*, 3912.

28. Admiraal, G.; Alink, M.; Altona, C.; Dijt, F. J.; van Garderen, C. J.; de Graaff, R. A. G.; Reedijk, J. *J. Am. Chem. Soc.* **1992**, *114*, 930.

29. Davies, M. S.; Berners-Price, S. J.; Hambley, T. W. *J. Am. Chem. Soc.* **1998**, *120*, 11380.

30. Wang, K.; Yu, Y. X.; Gao, G.-H.; Luo, G.-S. *J. Chem. Phys.* **2005**, *123*, 1.

31. Skylaris, C. K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. *Phys. Stat. Sol. B* **2006**, *243*, 973.

32. Raugei, S.; Gervasio, F. L.; Carloni, P. *Phys. Stat. Sol. B* **2006**, *243*, 2500.

33. Canfield, P.; Dahlbom, M. G; Hush, N. S.; Reimers, J. R. *J. Chem. Phys.* **2006,** *124*, 1.

34. McNamara, J. P.; Hillier, I. H. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2362.

35. Sha, S.; Roy, R. K. *J. Phys. Chem. B* **2007**, *111*, 9664.

36. Troisi, A.; Orlandi, G. *J. Phys. Chem. B*, **2002**, *106*, 2093.

37. van der Vaart, A.; Merz, Jr. K. M. *J. Chem. Phys.* **2002**, *116*, 7380.

38. Habza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247.

39. Foresman, J. B. *Exploring Chemistry with Electronic Structure Methods*, **1996**, *2nd ed.*, Gaussian, Inc.

40. Case, D. A.; III, T. E. C.; Tom Darden, H. G.; Luo, R.; Jr., K. M. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.

41. Zilberberg, I. L.; Avdeev, V. I.; Zhidomirov, G. M. *J. Mol. Struc. (Theochem).* **1997**, *418*, 73.

42. Pelmenschikov, A.; Zilberberg, I.; Leszczynski, J. *et al. Chem. Phys. Lett.* **1999**, *314*, 496.

43. Zeizinger, M.; Burda, J. V.; Leszczynski, J. *Phys. Chem. Chem. Phys.* **2004**, *6*, 3585.

44. Qin, H.; Li-Xin, Z.; Zhi-Qiang, Z. *Chin. J. Chem.* **2005**, *23*, 1355.

45. Kitaura, K.; Sawai, T.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *312*, 319.

46. Kitaura, K.; Sawai, T.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.

47. Takeshi, I.; Yuji, M.; Tatsuya, N. *et al. Chem. Phys. Lett.* **2006**, *427*, 159.

48. Frisch, M. J.; Trucks, G W.; Schlegel, H. B.; Scuseria, G E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, Jr. T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G; Rega, N.; Petersson, G A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.04; Gaussian, Inc., Pittsburgh PA, **2003**.

49. Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S. H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.*, **1992**, *63*, 751.

50. NDB ID: PD0005, PD0010, PD0023, PD0024, PD0030, PD0032, PD0033, PD0041, PD0068, PD0095, PD0105, PD0112, PD0145, PD0146, PD0155, PD0165, PD0192, PD0200, PD0219, PD0221, PD0222, PD0230, PD0231,

PD0251, PD0252, PD0259, PD0270, PD0286, PD0307, PD0309, PD0310, PD0312, PD0316, PD0319, PD0320, PD0338, PD0364, PD0365, PD0393, PD0424, PD0429, PD0504, PD0505, PDE003, PDE011, PDE012, PDE0128.

51. NDB ID: 169D, 1G5K, 1GJ1, 1IV6, 1J5N, 1KBD, 1KVH, 1LFU, 1LWA, 1N0K, 1N0O and 2KBD.
52. SYBYL, version 6.2; Tripos Inc.: Missouri, **1995**.
53. Raber, J.; Zhu, C.; Eriksson, L.A. *J. Phys. Chem. B* **2005**, *109*, 11006.
54. Rak, J.; Voityuk, A. A.; Marquez, A.; Rösch, N. *J. Phys. Chem. B* **2002**, *106*, 7919.
55. Guallar, V.; Douhal, A.; Moreno, M.; Lluch, J. M. *J. Am. Chem. Soc.* **1999**, *103*, 6251.
56. Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem. Eur. J.* **1999**, *5*, 3581.
57. Stephan, R.; Christel, M. M. *J. Comput. Chem.* **2007**, *28*, 1503.
58. Toru, M.; Yasuteru, S.; Kimihiko, H. *J. Phys. Chem. B* **2007**, *111*, 1176.
59. Alessandro, T.; Giorgio, O. *J. Phys. Chem. B* **2002**, *106*, 2093.
60. Mark, P. W.; Arturo, R.; James, A. P. *et al. J. Comput Chem.* **2006**, *27*, 491.
61. Huheey, J. E.; Keiter, E. A.; Keiter, R. L. *Inorganic chemistry: principles of structure and reactivity* **1993**, *4th ed.*, New York: HarperCollins College Publishers.
62. Bellon, S. F.; Coleman, J. H.; Lippard, S. J. *Biochemistry*, **1991**, *30*, 8026.
63. Pil, P. M.; Lippard, S. J. *Science*, **1992**, *256*, 234.

64. Moggs, J. G; Szymkowski, D. E.; Yamada, M.; Karran, P.; Wood, R. D. *Nucleic Acids Res.* **1997**, *25*, 480.

65. Lavery, R.; Sklenar, H. *CURVES 4.1*, **1992**, Institut de Biologie Physico-Chimique, Paris.

66. Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dynam.* **1988**, *6*, 63.

67. Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dynam.* **1989**, *6*, 655.

68. Wilkins, M. H. F.; Randall, J. T. *Biochim. et Biophys. Acta.* **1953**, *10*, 192.

69. Fichtinger-Schepman, A. M. J.; van der Veer, J. L.; den Hartog, J. H. J.; Lohman, P. H. M.; Reedijk, J. *Biochemistry*, **1985**, *24*, 707.

70. NDB ID: NMR: 169D, 1J5N, 1LFU; X-Ray: PD0024, PD0030, PD0033, PD0112, PD0222, PD0393, PD0504, PD0505, PDE003.